A Project Report on

Detecting Novelty Seeking from Online Travel Reviews: A Deep Learning Approach

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology In

Computer Science and Engineering (Artificial Intelligence and Machine Learning)

Submitted by

S.Niharika
(21H51A6653)
C.Baby Shalini
(21H51A66B6)
M.Shailaja
(21H51A66C8)

Under the esteemed guidance of

Dr.J.Sasi Bhanu
(Professor)

Department of Computer Science and Engineering(AI&ML)

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)

Approved by AICTE Affiliated to JNTUH NAAC Accredited with A+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2024- 2025

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD-501 401.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERNING(AIML)

CERTIFICATE

This is to certify that the Major Project Phase I report entitled "**Detecting Novelty Seeking from Online Travel Reviews** " being submitted by 21H51A6653, 21H51A66B6, 21H51A66C8 in partial fulfillment for the award of Bachelor of Technology in Computer Science and Engineering(AIML) a record of bonafide work carried out under my guidance and supervision. The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Dr.J.Sasi Bhanu

Professor

Dept. of CSE(AI & ML)

Dr.P.Sruthi

Professor and HOD

Dept. of CSE(AI & ML)

ACKNOWLEDGEMENT

With great pleasure I want to take this opportunity to express my gratitude to all the people who helped in making this Project a grand success.

We are grateful to **Dr.J.Sasi Bhanu, Professor** , Department of Computer Science and Engineering(AI&ML) for her valuable technical suggestions and during the execution of this Project Work.

We would like to thank **Dr.P.Sruthi,** Head of the Department of Computer Science and Engineering (AI&ML), who is the major driving forces to complete our project work successfully.

We would like to thank, **Dr.P.Ravikumar**,Dean-F&S, CMR College of Engineering and Technology, for his insights and expertise have been instrumental in shaping the direction and execution of this project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the Project work successfully.

We are highly indebted to Major **Dr.V A Narayana**, Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff **of Department of Computer Science and Engineering (AI&ML)** for sharing their knowledge with me.

We express our sincere thanks to **Sri. Ch. Gopal Reddy** garu, Secretary & Correspondent, CMR Group of Institutions, for his continuous care towards my achievements.

We express our sincere thanks to **Sri. Ch. ABHINAV Reddy** garu, CEO, CMR Group of Institutions, for his continuous care towards my achievements.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly or indirectly in completion of this project work.

S.Niharika 21H51A6653

C.Baby Shalini 21H51A66B6

M.Shailaja 21H51A66C8

CHAPTER NO	TITLE	PAGE NO
	LIST OF FIGURES	6
	LIST OF TABLE	7
	ABSTRACT	8
1	**INTRODUCTION**	
	1.1 Problem Statement	10
	1.2 Research Objective	10-11
	1.3 Project Scope and Limitations	11-13
2	**BACKGROUND WORK**	
	2.1 Manual Classification	
	2.1.1 Introduction	14
	2.1.2 Merits,Demerits and Challenges	14-15
	2.1.3 Implementation of Manual classification	16-17
3	**DISCUSSION AND RESULTS**	
	3.1 Comparision of Existing Solution	19
	3.2 Data Collection and Performance Metrics	20-21
4	**CONCLUSIONS**	
	4.1 Conclusion and Future Enhancement	23
6	**REFERENCES**	25

List of Figures

List of Tables

ABSTRACT

With the surge in user-generated travel content, understanding travelers' novelty-seeking behavior has become essential for personalization and targeted marketing in the tourism industry. Novelty-seeking refers to travelers' desire for unique, unfamiliar, and adventurous experiences, which can be observed in online reviews. This research proposes a deep learning-based approach to automatically detect novelty-seeking tendencies within online travel reviews. By leveraging natural language processing (NLP) and deep learning models such as Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT), the system analyzes textual patterns and sentiments indicative of novelty-seeking behavior. The dataset comprises diverse travel reviews from online platforms, preprocessed and annotated for novelty-seeking characteristics. Our model achieves high accuracy in classification, outperforming traditional machine learning techniques in detecting subtle linguistic cues associated with a traveler's interest in novel experiences. The findings offer practical implications for travel companies, enabling them to refine recommendation systems and better understand their customer base's psychographic profiles. This approach underscores the potential of AI in enhancing the tourism sector's ability to cater to evolving traveler preferences.

CHAPTER-1

INTRODUCTION

CHAPTER-1

INTRODUCTION

1.1 Problem Statement

In the tourism industry, understanding travelers' preferences is critical for providing personalized experiences. One significant aspect of traveler preferences is "novelty-seeking behavior," where travelers seek unique, unconventional, or adventurous experiences. However, detecting this behavioral trait in online travel reviews remains challenging due to the nuanced and subjective nature of language used in reviews, as well as the diverse ways in which novelty-seeking behavior is expressed.

Current recommendation systems primarily rely on explicit user ratings and tags, which may not capture deeper psychological traits like novelty-seeking. This gap limits travel companies' ability to tailor recommendations accurately to travelers seeking new experiences. Manual review of text data is impractical on a large scale, and traditional machine learning approaches struggle to capture subtle linguistic cues associated with novelty-seeking.

This project aims to address these challenges by developing a deep learning-based model that can automatically detect novelty-seeking tendencies from unstructured text in online travel reviews. Using advanced natural language processing techniques, the model will identify and classify novelty-seeking characteristics within reviews to enhance personalized travel recommendations, ultimately improving user satisfaction and engagement in the travel industry.

1.2 Research Objective

Travel online reviews is important experience related information for understanding an inherent personality trait, novelty seeking (NS), which influences tourism motivation and the choice of tourism destinations. Manual classification of these reviews is challenging due to their high volume and unstructured nature. This paper aims to develop a classification framework and deep learning model to overcome these limitations. A multi-dimensional classification framework was created for

NS personality trait that includes four dimensions synthesized from prior literature: relaxation seeking, experience seeking, arousal seeking and boredom alleviation. Based on 30 000 reviews from TripAdvisor we propose a deep learning model using Bidirectional Encoder Representations from Transformers (BERT)- Bidirectional Gated Recurrent Unit (BiGRU) to recognize NS automatically from the reviews. The classifier based on BERTBiGRU and NS multi-dimensional scales achieved precision and F1 scores of 93.4% and 93.3% respectively, showing that NS personality trait can be relatively accurately recognized. This study also demonstrates that the classifier based on multi-dimensional NS scales can produce satisfactory results using the deep learning model. The findings also indicate that the BERT- BiGRU model achieves the best effect compared to the same kind of deep learning models. Moreover, it proves that personality traits can be automatically identified from travel reviews based on computational techniques. For practical purposes, this study provides a comprehensive classification framework for NS, which can be used in marketing and recommendation systems operating in the tourism industry.

1.3 Project Scope and Limitations

Project Scope:

1. **Model Development**
    - Explore deep learning models suitable for text classification, such as:
        - Long Short-Term Memory (LSTM) networks
        - Bidirectional LSTM
        - Bidirectional Encoder Representations from Transformers (BERT)
    - Train the models to identify novelty-seeking patterns in travel reviews.
    - Compare with baseline machine learning methods (e.g., SVM, Naïve Bayes) for benchmarking.
2. **Feature Engineering**
    - Identify key features in the text data that may indicate novelty-seeking behavior, such as descriptive language, excitement-related vocabulary, or exploration mentions.
    - Implement NLP techniques (e.g., sentiment analysis, semantic analysis) to further extract insights into novelty-seeking characteristics.

3. **Evaluation and Testing**
    - Measure model performance using metrics like accuracy, F1-score, recall, and precision.
    - Conduct tests to assess model generalization on different subsets of travel review data (e.g., location, type of travel).
4. **Result Interpretation**
    - Analyze the model's output to identify trends in novelty-seeking preferences across demographics and locations.
    - Use model interpretations to understand the types of phrases and language that signal novelty-seeking in the reviews.
5. **Application Development**
    - Develop a proof-of-concept application for travel companies to test the model's classification on live or historical review data.
    - Integrate the model's output into recommendation systems to enhance personalized travel suggestions based on novelty-seeking tendencies.
6. **Documentation and Reporting**
    - Document the entire development process, including model selection, data processing, training, evaluation, and results.
    - Prepare reports on model insights and limitations, along with suggestions for further enhancement.
7. **Future Enhancements**

- Suggest improvements such as expanding the model to detect other psychographic profiles (e.g., relaxation-seeking, budget-conscious).
- Consider additional data sources like social media reviews to broaden the model's applicability and robustness.

Limitations:

- **Data Quality and Availability**

- **Data Bias**: Online reviews may have biases, as travelers who seek novelty may post disproportionately positive or negative reviews, skewing the dataset.

- **Incomplete or Inconsistent Data**: Online reviews can vary greatly in detail, style, and length, affecting the model's ability to accurately classify novelty-seeking tendencies across all review types.

- **Subjectivity in Annotation**

  - **Annotation Challenges**: Manually labeling novelty-seeking behavior in reviews can be subjective, as what is novel for one user may be familiar for another. This variability makes it challenging to achieve high-quality, consistent annotations.
  - **Dependence on Human Interpretation**: The dataset may require extensive human judgment to interpret novelty-seeking traits, which can introduce inconsistencies and reduce model accuracy.

- **Complexity of Novelty-Seeking Behavior**

  - **Linguistic Nuance**: Novelty-seeking can be challenging to identify due to nuanced language. Travelers may not always express novelty-seeking directly, making it difficult for the model to detect indirect or implicit cues.
  - **Varying Definitions of Novelty**: Novelty is subjective and varies based on culture, experience, and individual preferences, which may limit the model's ability to generalize across diverse audiences.

- **Model Generalization**

  - **Limited Transferability**: The model trained on specific platforms or datasets may not generalize well to other datasets due to differences in review style, language, and content.
  - **Overfitting Risks**: Deep learning models, particularly complex ones like BERT, can be prone to overfitting, especially with small or unbalanced datasets, leading to poor performance on new, unseen data.

CHAPTER-2

BACKGROUND WORK

CHAPTER-2

BACKGROUND WORK

2.1 Manual Classification of Reviews

2.1.1 Introduction

In Existing Manual classification of these reviews is challenging due to their high volume and unstructured nature reading above dataset user can understand about novelty seeking and it’s difficult to read all reviews to make decision Many existing deep learning algorithms such as CNN, LSTM are available but their prediction accuracy is not good enough. Existing methods for detecting novelty-seeking behavior in online travel reviews face major limitations due to the unstructured and vast amount of data involved. Manual classification is impractical and inconsistent, as human interpretation varies and is not scalable for the millions of reviews generated on travel platforms. Traditional machine learning models like SVM and Naïve Bayes have been applied but struggle with capturing the nuanced language patterns associated with novelty-seeking. These models rely on extensive feature engineering, which fails to fully account for the subtle and subjective expressions travelers use to describe novel experiences, leading to poor generalization and accuracy. Additionally, while some deep learning approaches, such as CNN and LSTM, have been attempted, they often lack the accuracy needed to reliably detect these complex behavioral cues, as they too struggle to interpret the rich diversity and context within user-generated text.

2.1.2 Merits, Demerits and Challenges

Merits

1. **Familiarity and Ease of Implementation**: Traditional machine learning models like SVM, Naïve Bayes, and Random Forests are well-known and straightforward to implement. They can be easily trained on labeled data without requiring extensive computational resources.
2. **Feature-Based Insights**: These models allow feature engineering, enabling some level of insight into specific linguistic markers associated with novelty-seeking, such as positive adjectives or mentions of unique activities, providing a basic level of interpretability.

3. **Foundational Results for Benchmarking**: While not ideal, traditional models and simpler deep learning approaches (CNN, LSTM) provide a useful benchmark for comparing newer, more complex models like transformers (e.g., BERT).

Demerits

1. **Poor Generalization**: Traditional models often fail to capture the nuanced and complex language patterns associated with novelty-seeking, leading to poor accuracy and generalization when exposed to new, diverse reviews.
2. **Reliance on Feature Engineering**: These models require extensive feature engineering to detect novelty-seeking tendencies. Manually crafted features often miss subtle cues, making the models less effective in understanding subjective and diverse expressions.
3. **Lower Accuracy in Deep Learning Models**: Early deep learning models like CNN and LSTM are often unable to accurately detect novelty-seeking behavior, as they struggle with context-heavy sentences and complex user-generated language, leading to suboptimal results.

Challenges

1. **Scalability and Real-Time Processing**: Processing large datasets of user-generated reviews is challenging for traditional models, which are less efficient in handling big data and often require manual feature extraction, making them impractical for real-time analysis.
2. **Subjectivity and Context in Reviews**: Novelty-seeking behavior is highly subjective, varying by user, context, and culture. Capturing these nuances in models is challenging without advanced, context-aware algorithms.
3. **Diversity in Expression**: Travelers express novelty-seeking differently, using slang, idioms, or region-specific language, making it difficult for traditional and early deep learning models to consistently detect these behaviors.

2.1.3 Implementation of Manual Classification of Reviews

Manual classification of reviews involves human reviewers reading and categorizing reviews based on the presence or absence of novelty-seeking behavior. Here's how the process would typically be implemented:

Steps for Manual Classification of Reviews

1. **Define Classification Criteria**
   - Establish clear guidelines to identify novelty-seeking behavior. For instance, novelty-seeking may include language describing unique experiences, adventure, exploration, unfamiliarity, or excitement.
   - Examples of novelty-seeking keywords: "adventurous," "unique," "unforgettable," "exotic," "first time," etc.
   - Non-novelty-seeking might include reviews focusing on comfort, familiarity, or routine.
2. **Sample Selection**
   - Select a representative sample of reviews, ensuring diversity in location, demographics, and review length.
   - To handle large datasets, take a stratified sample to maintain balance across different types of reviews (positive, negative, neutral).
3. **Reviewer Training**
   - Train reviewers on the classification criteria and provide examples to ensure consistency. Reviewers should understand novelty-seeking definitions to maintain uniformity.
   - Use practice reviews and discuss any ambiguous cases as a group to align on interpretation.
4. **Review and Labeling**
   - Each reviewer reads the reviews in the sample and labels them as "Novelty-Seeking" or "Not Novelty-Seeking" based on the established criteria.
   - Reviewers may also be asked to add notes on why they labeled each review, which can be helpful for later reference or clarification.

5. **Validation and Consensus Building**
    - After labeling, reviews can be cross-checked among reviewers to improve consistency.
    - For any disagreements, a consensus-building process can be used. Reviews that are difficult to categorize may be discussed among reviewers or involve a lead reviewer to make the final decision.
6. **Data Entry**
    - Record the labels in a structured format, such as a spreadsheet or database, for further analysis.
    - Labels may be binary (e.g., 1 for novelty-seeking, 0 for non-novelty-seeking) or involve additional categories if a more nuanced classification is needed.

Post-Classification Analysis

After manual classification, the labeled dataset can be used to train a model or generate insights. The dataset’s novelty-seeking tags become a valuable resource for machine learning models and can serve as ground truth for training and evaluating automated classifier

CHAPTER-3
RESULTS AND DISCUSSION

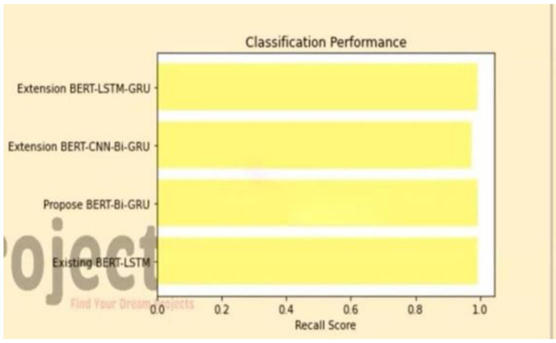
CHAPTER-3

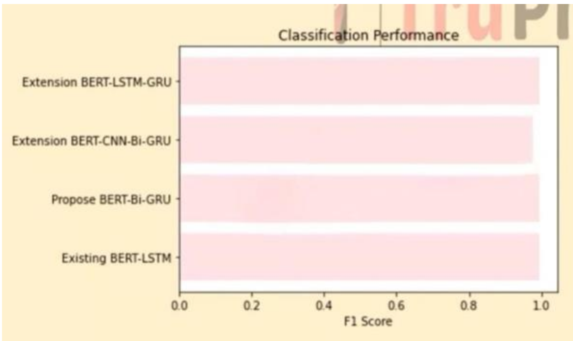
RESULTS AND DISCUSSION

3.1 Comparision of Existing Solutions

Comparison of existing solutions for detecting novelty-seeking behavior in online travel reviews, focusing on **manual classification**, **traditional machine learning models**, and **early deep learning models:**

Solution	Description	Advantages	Disadvantages	Best Suited For
Manual Classification	Human reviewers manually read and classify reviews as novelty-seeking or non-novelty-seeking based on predefined criteria.	- High interpretability and nuanced understanding. - Ideal for creating high-quality labeled data.	- Time-consuming and labor-intensive. - Inconsistent due to subjective interpretations. - Not scalable.	Small datasets or initial labeling for model training.
Traditional ML Models	Uses algorithms like **SVM**, **Naïve Bayes**, and **Random Forests**. Requires extensive feature engineering to capture novelty-seeking language patterns.	- Straightforward to implement. - Relatively fast on small datasets. - Allows some interpretability through features.	- Limited accuracy for subjective, nuanced behaviors. - Requires manual feature engineering. - Poor generalization.	Small, structured datasets with simple text classification.
Early Deep Learning Models	Models like **CNN** and **LSTM** are used to learn complex patterns in review text. CNNs identify local patterns, while LSTMs capture sequential relationships in the text.	- Can capture deeper language patterns than traditional models. - No manual feature engineering required.	- Often inadequate accuracy due to lack of deep context awareness. - Computationally intensive. - Prone to overfitting on small datasets.	Medium-sized datasets with less need for high contextual accuracy.





3.1.1 Classification Performance

3.2 Data Collection and Performance Metrics

1. Source Identification

- **Online Travel Platforms**: Collect reviews from popular travel platforms such as TripAdvisor, Yelp, Expedia, or Google Reviews. These platforms have a vast array of user-generated content regarding different travel experiences.

2. Data Extraction

- **Web Scraping**: Use web scraping tools (e.g., Beautiful Soup, Scrapy) or APIs (if available) to extract reviews. Focus on key fields such as:
  - Review text
  - Ratings (if available)
  - User demographics (location, type of traveler)
  - Date of the review
- **Existing Datasets**: Explore existing datasets available in research communities or Kaggle that may already contain labeled travel reviews.

3. Data Preprocessing

- **Cleaning**: Remove any irrelevant data (HTML tags, special characters) and handle missing values.
- **Normalization**: Standardize text by converting to lowercase, removing stop words, and applying stemming or lemmatization.
- **Annotation**: If using manual classification, ensure that reviews are annotated for novelty-seeking behavior based on the established criteria.

4. Data Storage

- Store the collected data in a structured format, such as CSV files or a database, with columns for the review text, rating, novelty-seeking label, and any other relevant information.

Performance Metrics

To evaluate the effectiveness of the model developed for detecting novelty-seeking behavior, the following performance metrics should be used:

1. Accuracy

- Measures the proportion of correctly classified instances (both novelty-seeking and non-novelty-seeking) out of the total instances.

  Accuracy=Total InstancesTrue Positives+True Negatives

2. Precision

- Measures the proportion of true positive predictions among all positive predictions made by the model.

  Precision=True Positives+False PositivesTrue Positives

3. Recall (Sensitivity)

- Measures the proportion of true positive predictions among all actual positive instances in the dataset.

  Recall=True Positives+False NegativesTrue Positives.

4. F1-Score

- The harmonic mean of precision and recall, providing a balance between the two metrics.

  F1-Score=2×Precision+RecallPrecision×Recall

5. Confusion Matrix

- A table used to evaluate the performance of the classification model. It shows the true positive, true negative, false positive, and false negative counts, allowing for a more granular view of model performance.

6. ROC-AUC Score

- The Area Under the Receiver Operating Characteristic curve (ROC-AUC) measures the model's ability to distinguish between classes. A score of 1 indicates perfect classification, while a score of 0.5 indicates random guessing.

CHAPTER-4

CONCLUSION

CHAPTER-4
CONCLUSION

4.1 Conclusion

Detecting novelty-seeking behavior in online travel reviews is a complex but crucial task for enhancing personalized travel experiences. Existing methods, including manual classification, traditional machine learning models, and early deep learning approaches, each have their strengths and limitations. While manual classification provides nuanced insights, it is impractical for large datasets. Traditional machine learning models offer quick implementations but struggle with the subjective nature of reviews, and early deep learning models, while more capable of capturing linguistic nuances, often lack the necessary context awareness for accurate detection.

The proposed deep learning approach aims to overcome these challenges by leveraging advanced natural language processing techniques to automatically classify novelty-seeking behavior in travel reviews. By utilizing large, labeled datasets, the model can learn to identify and interpret the subtle cues associated with novelty-seeking, leading to improved accuracy and scalability in detection.

Ultimately, implementing a robust solution for detecting novelty-seeking behavior can significantly benefit travel companies by enabling more personalized recommendations and enhancing user satisfaction. As the travel industry continues to evolve, integrating advanced technologies into the review analysis process will be vital for understanding and catering to the diverse preferences of travelers.

CHAPTER-5

REFERENCES

CHAPTER-5

REFERENCES

5.1 References

Papers

1. Zhang, Y., & Wallace, B. (2015). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *arXiv preprint arXiv:1510.03820*.
2. Yang, Z., et al. (2016). Hierarchical Attention Networks for Document Classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Books

1. Zuckerman, M. (1994). *Behavioral Expressions and Biosocial Bases of Sensation Seeking*. Cambridge University Press.
   Link to Book
2. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human-Centered Informatics, 5(1), 1-167.
   Link to Lecture

Links

1. Vermeulen, I. E., & Seegers, A. (2009). Tired of the Same Old Places? The Influence of Novelty-Seeking on the Relationship Between Online Reviews and Tourists' Decisions. *Tourism Management*.
   Link to Journal
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
   Link to Paper