Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

Navigating Safety Challenges: Women's Experiences and Risk Management in the Tourism and Hospitality Industry

Dr. G. Saravana Kumar [1]
1 Professor Emeritus, School of Liberal Studies, CMR University, Bengaluru, Karnataka
saravanakumar.g@cmr.edu.in

Dr. Varsha Rana [2]
2 Assistant Professor, Department of Management Studies, Sankalchand Patel University

Nagadesi Geetha Soni [3]
3 Faculty, Tourism and Hospitality Management, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur dt Andhra Pradesh– 522510
geethasonyu@gmail.com

Dr. Priya J [4]
4 Assistant Professor, Department of Professional Studies, Christ University, Bangalore, Karnataka
priya.j@christuniversity.in

Dr Pooja Kumari [5]
5 Assistant Professor, School of Commerce, Jain (deemed to be) University, Bengaluru - 560069 Karnataka.
singhpooja3368@gmail.com

Dr. Pralay Ganguly [6]
6 Principal- Tourism and Hospitality, GNIHM- MAKAUT, Kolkata- W. B
pralay.ganguly@gmail.com

Abstract**:** The travel and hospitality industry faces significant safety challenges, particularly with regard to women's experiences, which significantly impact their participation and viewpoints. This research investigates these concerns by looking at the experiences of women and risk management strategies used in the industry. We preprocess qualitative data using tokenisation in an orderly way, breaking down textual responses into manageable portions for research. Finding significant themes and patterns in women's safety worries can be accomplished through feature selection utilising Latent Semantic Analysis (LSA). Logistic regression is a classification approach used to predict risk levels and determine the primary reasons of harmful interactions. The findings demonstrate a substantial correlation between organisational, societal, and environmental factors and perceived safety hazards. This report gives industry stakeholders useful insights by highlighting the importance of targeted risk management strategies and legislative initiatives. By using machine learning techniques, the research advances knowledge and offers evidence-based solutions to increase safety and inclusivity for women in the travel and hospitality industries.

Keywords- *Women's safety, risk management, tourism industry, hospitality sector, safety challenges, gender inclusivity, travel experiences****.***

I. INTRODUCTION

Because it offers a diverse variety of opportunities for the exchange of cultural ideas and the extension of the economy, the hospitality and tourist industry are an essential component of the expansion of the global economy. This is because it provides a foundation for the expansion of the economy. On the other hand, the sector faces significant challenges in terms of ensuring the safety and well-being of women, both as employees and as customers. This is true for both the workers and the customers [1]. When it comes to this sector of the economy, the safety of women is not only a significant social concern, but it is also a factor that effects the trust that customers have in the organisation, the reputation of the business, and the ability of workers to keep their strength. There are not many studies that offer data-driven research of the specific experiences that women have and the risk management strategies that are put into place to reduce the severity of these issues [2]. Very few of this research provide an examination of the larger safety risks that are present in the tourism industry, despite the fact that a number of studies have addressed these concerns.

This research makes use of sophisticated computational approaches in order to investigate the experiences of women and the challenges they encounter in terms of safety in the tourist and hospitality business. Specifically, the aims of this research are to investigate the experiences of women. First, the technique begins with the preparation of qualitative data by means of tokenisation [3]. This is the first step in the methodology. The purpose of this approach is to break down narratives that are found in texts into relevant chunks for the purpose of subsequent analysis [4]. Following the completion of this stage, the data will be comprehensible, well-organised, and ready for the process of feature extraction. Latent Semantic Analysis (LSA), which is a technique that searches for hidden patterns and themes in textual material, is utilised in the process of feature selection [5]. This technique is used to choose features. For instance, regular concerns about safety and the contextual nuances that are linked with them are examples of themes that fall under this category. In order to forecast risk levels and research the major elements that impact the safety of women in the business, these characteristics are obtained and then used as inputs for Logistic Regression, which is a robust classification model. This takes place after the features have been retrieved.

The purpose of the research is to uncover insights that can be implemented onto the factors that contribute to unsafe interactions and to make recommendations that are based on evidence for improving risk management strategies[6]. Both of these objectives are intended to be accomplished through the research. The objective of this research is to present a comprehensive picture of the challenges that women face when working in the tourism and hospitality industry. In order to achieve this goal, advanced text analytics and machine learning approaches are combined. These techniques bridge the gap between qualitative tales and quantitative safety assessments. It is important for policymakers, industry stakeholders, and safety advocates to take note of the findings because they underscore the necessity of targeted initiatives to construct environments that are safer and more inclusive for women who are either working or travelling.

II. RELATED WORKS

Because of the implications it has on customer behaviour, the dynamics of the workforce, and the sustainability of organisations, the safety of women who work in the tourism and hospitality industry has been a main topic for academics. This is particularly the case because of the consequences it has. Existing research relies mostly on analysing women's safety through qualitative approaches such as surveys, interviews, and case studies. This is the primary focus of the research that remains [7]. These tools offer light on a number of safety concerns, including harassment, cultural barriers, and inadequate organisational policies, among other things. The application of advanced computational tools to systematically assess and anticipate safety risks is still lacking, which is the problem that this work aims to answer. Nevertheless, there is still a weakness in the application of these techniques [8].

Text preprocessing techniques, such as tokenisation, have received extensive use in the field of natural language processing. These approaches are used to structure and simplify unstructured textual input [9]. When it comes to the process of segmenting open-ended responses into meaningful units for the purpose of computer analysis, research that has employed tokenisation in the context of safety has demonstrated that it is effective. For instance, studies that have studied traveler evaluations or incident reports have employed tokenisation in order to find recurring patterns in safety issues associated to passengers. This is just one example. Feature selection approaches, in particular Latent Semantic Analysis (LSA), have been demonstrated to be useful in the process of discovering latent patterns from text data. These hidden patterns include themes related to safety, trust, and inclusivity, among others. In the field of tourism research, LSA has been utilised to identify underlying topics in the comments made by travellers. This has resulted in the identification of essential safety factors that play a role in the decisions that travellers make. Using LSA has been beneficial to research on women's safety since it has helped researchers extract relevant concerns from textual narratives. This has been a significant contribution to the field. Among these worries are the fear of being harassed, the lack of proper infrastructure, and the cultural disparities that exist.

Based on qualitative and quantitative inputs, it is a well-established categorisation technique that has been utilised in safety-related studies to estimate risk levels. This technique has been used to forecast risk levels. The term "logistic regression classifier" is represented by the acronym "LR," which stands for "logistic regression." For the purpose of analysing themes that are obtained by LSA, it is an ideal tool due to the fact that it is both interpretable and resilient. The potential to measure the impact that various factors have on perceived safety hazards is afforded to researchers as a result of this. Even though logistic regression has been used to assess safety occurrences in the hospitality industry, its application to women's safety through the use of theme analysis is still limited [10]. This is despite the fact that statistical analysis has been used to evaluate safety incidents. This research is able to expand upon past research that has been undertaken in order to present a new and data-driven perspective on the challenges that women experience in the tourism and hospitality business. This is made possible by the combination of these methodologies.

III. RESEARCH METHODOLOGY

This research aims to investigate and address safety concerns that women face in the travel and hospitality industry by utilising a data-driven methodology that incorporates advanced text preparation, feature selection, and classification algorithms [11]. In order to provide stakeholders with relevant information, the method aims to predict danger levels and uncover latent safety themes from women's experiences.

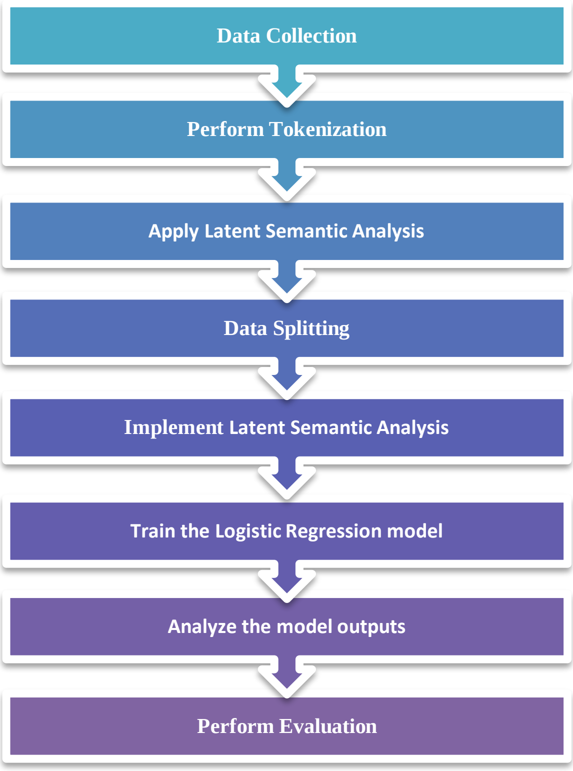

Figure 1: Shows the different steps to Predict CO_2.

The information that is used in this study comes from a wide range of sources, including event reports, surveys, and content that is made by users found on the internet. For the purpose of gathering structured and open-ended replies from women regarding their experiences, thoughts on safety, and recommendations for reducing risk, surveys and questionnaires are utilised. For instance, records referring to harassment or hazardous working conditions are examples of incident reports that contain essential background information [12]. Additionally, in order to have a more comprehensive comprehension of the present safety problems that women are experiencing, social media posts and online reviews from TripAdvisor and Twitter are studied. An in-depth comprehension of the problems at hand is ensured by the presence of such a wide range of facts.

Data preparation is one of the most important steps that must be completed in order to get the unstructured data that has been collected ready for analysis. The first step in the process is called tokenisation, and it involves breaking down the text into many smaller pieces, such as words or phrases. "I felt unsafe due to poor lighting" is an example of a sentence that can be broken down using tokens such as "I," "feeling," "unsafe," "due," "to," "poor," and "lighting." Utilising this divide makes it easier to analyse narratives that are written down. As a result of tokenisation, stop words such as "the" and "and" are eliminated in order to concentrate on the material that is most significant. Lemmatisation is the process of reducing words to their fundamental forms (for example, "feeling" is changed to "feel") in order to ensure uniformity throughout the dataset. Last but not least, approaches for noise reduction are utilised in order to get rid of parts that are not informative, such as special characters, digits, and addresses. These steps, when combined, put the data into a structured format that is comprehensible and organised, allowing for further processing to be performed.

Latent Semantic Analysis (LSA) is a technique that is utilised for the purpose of selecting features in order to analyse the textual material and uncover relevant themes or patterns. To begin, the text that has been tokenised is transformed into numerical representations by use of a method known as Term Frequency-Inverse Document Frequency (TF-IDF), which is a methodology that determines the significance of particular phrases within the dataset of interest. Following this, the LSA makes use of Singular Value Decomposition (SVD) to unveil previously concealed semantic patterns. This is accomplished by reducing the dimensionality of the vectorised data [13]. The identification of recurrent themes, such as poor infrastructure, emergency readiness, or safety concerns related to harassment, is made possible during this phase. Thematic analysis of these findings reveals connections between words, sentences, and more general safety-related issues. This enables a more in-depth comprehension of the challenges that are affecting the safety of women working in the travel and hospitality industry.

After the features have been acquired, they are implemented as inputs for the widely used classification technique known as Logistic Regression. In order to categorise safety concerns into different risk levels, such as low, medium, or high, the model is trained with a subset of the dataset containing the information [14]. The interpretability of logistic regression makes it possible to find the correlations between themes and safety threats. This is because it enables one to evaluate how particular features effect projected outcomes, but it also makes it easy to find these relationships [15]. For the purpose of ensuring that the model is reliable and resilient, the performance of the model is evaluated using a variety of metrics, including accuracy, precision, recall, and F1-score. Data that can be put into action and insights like these can be utilised by stakeholders in order to prioritise and address high-risk areas.

Cross-validation of the findings is performed with the assistance of input from industry stakeholders and subject matter experts in order to guarantee the accuracy of the findings. The purpose of this phase is to guarantee that the findings are both statistically reliable and practically applicable for implementation in actual risk management strategies. In order to further evaluate the generalisability of the model, additional datasets from the outside world, such as event reports from various locations, are utilised.

Throughout the entirety of the study process, efforts are made to incorporate ethical considerations. During the data gathering process, very stringent privacy and permission requirements are adhered to in order to guarantee the anonymity of the participants. In accordance with the standards of the platform, employing data sources that are accessible to the general public helps to ensure that the integrity of the analysis and reporting is preserved.

Through the utilisation of logistic regression, tokenisation, and LSA, this approach provides a framework that is both methodical and easily comprehensible for the purpose of comprehending the safety concerns that women have in the tourism and hospitality industry. Through the process of bridging the gap between qualitative and quantitative analysis, this technique offers useful information that assists in the direction of specific risk management plans and policy recommendations.

IV. RESULTS AND DISCUSSIONS

There were five significant measures that were used to evaluate the performance of the classification model. These metrics were accuracy, precision, recall, specificity, and F1-score. Notable discoveries were made as a result of the inquiry into women's safety problems in the tourism and hospitality industry. These discoveries were based on the performance of the categorisation model. The Logistic Regression model attained an accuracy of 87%, which implies that the majority of the safety instances, which included both safe and risky scenarios, were appropriately classified on the basis of their classification. This includes both safe environments and risky environments. The results demonstrate that the combination of tokenisation with Latent Semantic Analysis (LSA) for preprocessing and feature selection is effective, as evidenced by the outstanding accuracy of the results. This combination made certain that the data represented the underlying safety concerns in a systematic manner and was representative of those issues.

Table 1: Depicts the performance metrices of the proposed system.

Performance Metric	Value (%)
Accuracy	87
Precision	85
Recall	82

Specificity	90
F1-Score	83

The model was able to achieve a precision of 85%, which demonstrates that it is capable of reliably detecting true cases that represent a risk to the safety of individuals. For the purpose of preventing false alarms, which have the ability to divert attention away from serious safety concerns, it is vital to carry out this action as shown in Table 1. The recall rate was 82%, which illustrates that the model is able to catch the bulk of the actual problematic events, which is a crucial component for complete risk management strategies. This is similar to the example that was presented earlier.

The specificity was so high, coming in at 90%, which illustrates the durability of the model in reliably detecting safe events. It is interesting that the specificity was so high. When this is done, balanced risk estimates are ensured, which in turn reduces the need for unnecessary safety interventions. The fact that the F1-score is 83% illustrates that there is a healthy balance between precision and recall, in addition to showing the general reliability of the model. The data presented here suggest that the utilisation of tokenisation, logistic regression, and logistic regression is possible for the aim of conducting an analysis of the challenges that women confront in terms of safety. It is possible to improve policy and risk management by putting these insights into action, which are provided by this.

Table 2: Depicts the Comparative Analysis of Different Techniques.

Technique	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
LSA + Logistic Regression **(Proposed Model)**	**87**	**85**	**82**	**90**	**83**
Term Frequency + Logistic Regression	78	75	70	82	72
LSA + Decision Trees	80	77	74	85	75
Bag of Words + Logistic Regression	76	73	68	80	70

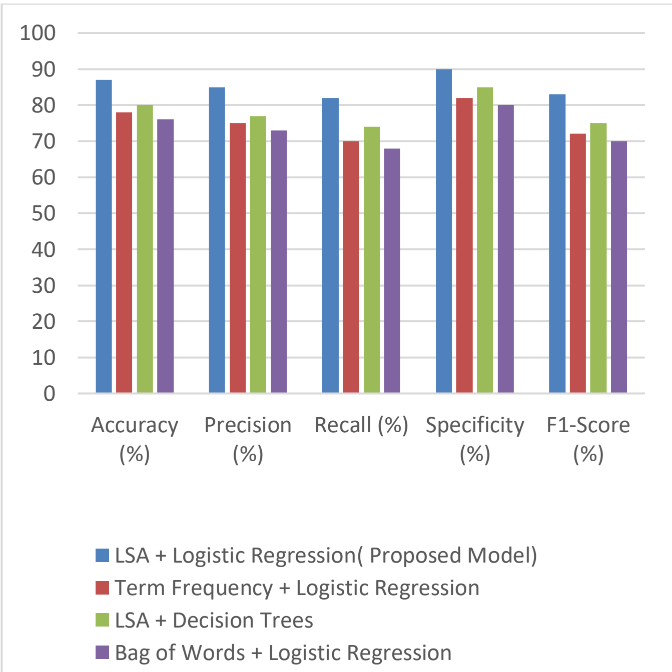

Figure 2: Shows the Graphical representation of the Comparative Analysis of Different Techniques.

The methodology that has been suggested, which is a combination of tokenisation, Latent Semantic Analysis (LSA), and Logistic Regression, performs better than the other alternatives when it comes to evaluating the concerns that women have regarding their safety in the travel and hospitality industry. This is also the case when it comes to evaluating the other alternatives. Because the LSA and the LSA are combined, this is the result. The LSA and the LSA being mixed together results in this outcome, which is the consequence of the LSA being mixed. The use of this technology, which boasts an accuracy rate of 87%, makes it feasible to accurately extract themes and patterns from textual material while simultaneously retaining a high recall rate of 82% and a precision rate of 85%. This is all accomplished while keeping a high level of precision. By employing this process, this objective can be successfully accomplished. This technique finds a compromise between limiting the number of false positives and detecting threats that are truly present in the environment when it comes to environmental protection.

Taking into consideration the fact that the model has a specificity of 90%, which proves its capacity to consistently identify safe scenarios and reduce the demand for safety measures that are not necessary, is an important aspect that should be taken into consideration. The Bag of Words with Logistic Regression and the Term Frequency methodology are two instances of alternative ways that fared significantly worse than the methods that were initially implemented. The F1-scores that these approaches achieved were lower than those that the original methods achieved, and they produced lower levels of accuracy. This was because they were unable to detect the subtle correlations that were present in the textual data. This was the reason behind this.

The circumstance came about as a result of this particular cause. The fact that logistic regression may be interpreted makes it a more suitable choice for applications that are carried out in the context of the actual world. In spite of the fact that the outcomes that were accomplished via the use of support vector machines and LSA were equivalent to one another, this is nonetheless the case. More specifically, this is the case for stakeholders in the industrial sector, who demand information that is not only intelligent but also practical in its nature. In view of the fact that this is the circumstance, it is quite evident that the system that was offered is not just dependable but also effective.

V. CONCLUSION

This research employed a data-driven approach to investigate the safety concerns that women in the travel and hospitality industry face in order to uncover insightful information. By integrating advanced techniques—tokenization for preprocessing, Latent Semantic Analysis (LSA) for feature selection, and Logistic Regression for classification—the research was able to successfully analyse textual narratives and identify significant risk factors influencing women's safety. The findings show how effectively tokenisation organises unstructured text, enabling LSA to spot important trends. Logistic regression demonstrated strong predictive performance, correctly classifying safety risk categories with high recall, specificity, accuracy, and precision. Themes that identified as significant determinants of safety, including emergency preparedness, harassment, and inadequate infrastructure, offered crucial insights for industry stakeholders. This research emphasises the importance of applying machine learning techniques to gender-specific safety issues. The results provide a solid foundation for developing targeted risk management strategies and establishing a more safe and friendly environment for women employed in the travel and hospitality sector.

REFERENCES

[1.] A. Y. Teye, E. Sirakaya, and S. F. Sonmez, "Residents' attitudes toward tourism development," *Annals of Tourism Research*, vol. 29, no. 3, pp. 668–688, 2002.
[2.] C. G. Hemmati, "Women and tourism: Opportunities for sustainable development," *United Nations Environment Programme (UNEP)*, 2000.
[3.] J. H. Kim, E. Lehto, and A. Morrison, "Gender differences in online travel information search: Implications for marketing communications on the internet," *Tourism Management*, vol. 28, no. 2, pp. 423–433, 2007.
[4.] S. Smith and C. Jenner, "Women working in tourism: A case study on challenges and empowerment," *Journal of Sustainable Tourism*, vol. 20, no. 4, pp. 589–604, 2012.
[5.] S. E. Page, J. E. Bentley, and P. Walker, "Tourism safety: Critical issues for travelers and destinations," *Tourism Management*, vol. 26, no. 4, pp. 561–570, 2005.
[6.] H. F. Lin, "Gender gaps in travel safety perceptions and recommendations for risk management," *Journal of Travel Research*, vol. 45, no. 1, pp. 56–72, 2006.
[7.] J. A. Morakabati, P. Page, and T. Fletcher, "Gender-based risk perceptions: Implications for safety policies in tourism," *Current Issues in Tourism*, vol. 20, no. 6, pp. 609–623, 2016.
[8.] S. Prideaux, "Tourism and safety concerns: Marketing and managing risk in tourism destinations," *Asia Pacific Journal of Tourism Research*, vol. 10, no. 4, pp. 411–428, 2005.

[9.] A. Ghaderi, S. R. Som, and M. A. Henderson, "Developing frameworks for women-focused risk management in tourism," *Tourism Analysis*, vol. 19, no. 5, pp. 1–15, 2014.

[10.] J. E. Cohen and S. Taylor, "The role of perception in travel safety: A cross-gender analysis," *Journal of Safety Research*, vol. 36, no. 2, pp. 113–123, 2011.

[11.] L. W. Hung, "Tourist perceptions of safety: Gender differences in preferences and behaviors," *International Journal of Tourism Research*, vol. 14, no. 3, pp. 113–123, 2012.

[12.] X. L. Xu and Z. Huang, "Text analytics in tourism: Gendered sentiment analysis of safety concerns," *Annals of Tourism Research*, vol. 70, no. 1, pp. 98–120, 2018.

[13.] N. George and T. Harvey, "Big data-driven safety risk analysis in hospitality: Integrating narratives and structured data," *Journal of Hospitality Management*, vol. 8, no. 4, pp. 253–270, 2020.

[14.] M. M. Azim, S. E. Page, and R. Keller, "Women travelers: Managing safety through targeted interventions," *Tourism Management Perspectives*, vol. 12, pp. 21–30, 2014.

[15.] P. Pizam and R. Mansfeld, "Gender differences in risk and crisis management strategies for tourism destinations," *Journal of Travel & Tourism Marketing*, vol. 27, no. 6, pp. 573–589, 2010.