

# A Smart Dashboard Framework for Urban Tourism Risk Analysis Using Deep Learning and Machine Learning

Adona Kulathinal Joseph  
MSc Computer Science

School of Computing, Engineering and Built Environment  
Ulster University, Belfast, United Kingdom  
Kulathinal\_Joseph-A@ulster.ac.uk

**Abstract** — This research presents an intelligent, data-driven method for tourism risk evaluation in Indian cities by integrating crime statistics, hotel reviews, and tourist opinions. An optimized BERT model accurately classifies user reviews into safety-related categories, whereas XGBoost is used for predicting crime trends and predicting city-level safety scores. Cluster analysis reveals distinctive city risk profiles based on crime rates and review sentiments. A strong negative relationship between tourist satisfaction and crime measures emphasizes the importance of safety in destination appeal. Evidence for these findings is presented through an interactive Power BI dashboard enabling real-time filtering, trend analysis, and predictive analysis. The system enables data-driven decision-making by travellers, policymakers, and city planners for safety assurance, demonstrating the synergy provided by combining AI and public data for tourism intelligence.

**Keywords** — Crime analytics, Tourism safety, BERT, Cluster analysis, predictive modelling, XGBoost, sentiment analysis, cluster analysis, Power BI, urban intelligence, Safety Dashboard.

## I. INTRODUCTION

The hospitality and tourism sector in India is increasingly being affected by rising crime rates, which pose significant risks to both tourist and local communities. While platforms such as TripAdvisor and Airbnb offer user reviews, they typically lack integration with comprehensive, real-time crime data. As a result, tourists often make travel decisions without clear understanding of local safety conditions. There is an evident connection between the tourism and victimisation of tourists. This is due to significant factors that influence the tourist choice [1].

This research addresses the critical gap in travel safety information by proposing a data-driven framework that combines tourism and accommodation review data with official crime reports, and temporal-spatial trends to analyse correlation between crime patterns and tourism activity. Using the merged dataset of tourism, accommodation and crime, the study applies data analysis, machine learning and deep learning techniques to uncover key insights and patterns and presents the results through an interactive dashboard.

Several issues emerged because of tourism industry's explosive growth. Overpopulation at tourist destinations affects the environment and leads to safety concerns [2]. Unlike conventional travel advisory platforms, this system allows users to filter data by crime type, city and time, thereby offering a more nuanced and actionable view of destination safety. The objective is to empower travellers to make informed decisions based on authentic, data-backed insights, while also offering a tool that can support tourism boards, local authorities, and smart city initiatives in enhancing public safety. This project not only contributes to safer and more informed tourism but also represents a novel step in integrating heterogeneous datasets for public benefits through intuitive visual analytics.

## II. LITERATURE REVIEW

The hospitality and tourism sector holds a vital position within the economy. Tourism serves as a driving force for economic growth, a stabilizing factor, and a unifying element in society by establishing direct and indirect links with various industries such as agriculture, manufacturing, transportation, trade, and others [3]. Tourism is a widespread and significant phenomenon, particularly for developing regions, where it is often seen as beneficial. However, recent studies highlight a different perspective. The temporary influx of large numbers of visitors into a concentrated area introduces outsiders to local communities, which can sometimes lead to challenges, including an increase in crime [4]. Most of the research bulk in tourism and crime is targeted at incidents of crimes by locals against tourists, which referred to as tourist-oriented crime, and has been well charted in quantitative and qualitative research. In contrast, it is opposite directions, such as tourist crimes against locals and those offenses with fellow tourists, are considered largely uncharted areas of concern within the academic literature. It also includes, within the broader context of the study of crime and tourism, crimes among locals that are related to tourism but do not directly involve visitors [5]. The connection between tourism, safety, and security has been examined by both tourism researchers and criminologists. Crime rates in tourist resorts tend to be elevated, particularly during peak travel seasons [6]. In 2016, India's tourism sector contributed approximately 208.9 billion U.S. dollars to the country's GDP, making it the second-largest contributor to GDP from tourism in the Asia-Pacific region, following China. Given its significant economic impact, tourism plays a crucial role in economic growth. Studies highlight that various criminal activities and terrorism influence tourists' decisions when choosing a destination. Violent crimes, such as murder and theft, have been identified as key factors affecting tourist preferences [7]. Academic studies have definitively established the link between tourism and crime. Since high criminal levels lead fewer people and lower hotel occupancy, travel destinations and lodging businesses are negatively affected by crime [8]. Early data shows that during the most travelled seasons, property crime rates in Miami surged. Research often points to tourist destinations as hotspots for criminal activities; for example, studies show how perceived and actual safety risks affect consumers' choices. Similar patterns were observed in Mexico, there was a correlation of tourist and crime. Research carried out in locations such as Hawaii and Tonga has also shown that unexpected consequence of tourism growth is a rise in crime rates [9]. Furthermore, present tourism sites including TripAdvisor and Airbnb offer useful reviews but do not include statistics on crime rates, therefore restricting the extent of information open to users. Different studies have investigated visualizing and analysing crime patterns. For example, research investigates how visitors affect crime rates, thereby exploring the link between travel and

crime. The author explored using empirical techniques how many different crimes across 103 Italian regions in 2005 were impacted by tourist numbers given sociodemographic and economic variables. It accounts for crime spillover effects, and it uses two geographic models: the spatial error and the spatial lag model. This study also investigates the effect of multiple tourist destinations urban, mountain and coastal areas on crime rates. An evaluation is done of the social costs of crime associated with the rise in tourism [10]. A case study of St. Lucia identified that among the property-related crimes against tourists, theft and burglary were relatively common, especially in those areas with significant tourism development. It reiterated that even though the citizens are more often victims of crime, incidents in tourist areas may have consequences for the tourism industry [11]. Another research investigates how the demand for tourists in Indian states is affected by crime and terrorist activity. and the results validate that violent and non-violent crimes have differing effects on India's tourism demand. Tourist arrivals decrease by 5.2% and 1.6% for every 10% rise in violent and non-violent crimes, respectively [12]. This project attempts to examine the effects of crime at different spatial range. To achieve this purpose, to analyse many data sources, such as travel evaluations, housing ratings, and local crime statistics in India. The findings of this study help us to understand the effects of crime and tourism industry and provide insightful suggestions for effective management of tourism destinations to ensure the safety of both traveller's and hosting communities.

### III. METHODOLOGY

The methodical procedures used to examine the effect of crime on Indian tourism are described in this section. The composition of methodology means data collection, preprocessing of the data analysis development of dashboard, and implementation of respective features to get comprehensive insights regarding the effect of crime on tourism in India.

#### A. Data Acquisition and Preprocessing

The collection of structured and unstructured data from different sources such as The India Crime Dataset, OYO Hotel Reviews and Indian Place to Visit Review dataset from Kaggle dataset repository. The crime dataset gives comprehensive insight into criminal activities in several Indian cities from 2020 to 2024, with a breakdown by type of crime, date and time of occurrence of crime, victim description such as age and gender of the victim. The OYO Hotel review dataset contain ratings and review for different hotel, beneficial in assessing how the user sentiments affect visitors view on safety [13], thus helping the business in enhancing the services and improving customer satisfaction. The tourism dataset encompasses city-based information, reviews and ratings left by visitors at various locations. Data preprocessing was performed to provide data quality, homogeneity and preparation for subsequent analysis [14], employing R with relevant libraries such as dplyr, data.table and tidytext. It comprised several steps such as cleansing, feature engineering, and normalization of three fundamental datasets. Irrelevant columns were identified and removed from each dataset to improve model performance.

Important preprocessing task was conducting sentiment analysis with the use of Bidirectional Encoder Representations from Transformers (BERT), employs a transformer model to completely

grasp the context of a word in a sentence in both directions, thus being more successful in classification tasks compared to conventional Natural Language Processing (NLP) models [15]. Using meticulously fine-tuned BERT models, text reviews collected from tourism as well as hotel datasets were labelled into four sentiments. To ensure consistency during merging, city names were standardized using string normalization techniques. A left join was applied based on city names as the key to combine the cleaned Crime, Tourism, and Hotel datasets. The approach enabled the retention of all the cities in the crime dataset.

#### B. Descriptive Statistics and Exploratory Data Analysis

Descriptive and Exploratory Data Analysis (EDA) were performed to check preliminary analysis on data to identify patterns, identify anomalies and to check assumptions using summary statistics and visualisation [16]. Descriptive statistics were obtained using the summary () function to have an initial view of numerical variables. Visualization including histogram, boxplots, bar charts were used for understanding data distribution. Use of boxplot is essential to identify any outliers of the primary numeric fields. Heatmap of correlation by using the ggcormplot package used to visualize an image of pairwise correlation strength and direction between variables. These tests permitted the simple comprehension of the data, shape, and for variable interactions.

#### C. Normality Test

To verify the dataset were normally distributed, a complete normality analysis was conducted. The skewness and kurtosis descriptive statistics were calculated for all the numeric variables to compare the symmetry and peakiness of the distribution of data respectively. The Anderson–Darling test was selected as a more suitable alternative to the Shapiro–Wilk test for determining the normality of continuous variables. The Anderson–Darling test is more sensitive towards the tails of the distribution [17], best suited for large data sets where small non-normalities are common but not consequential. The test was applied to all numeric fields of the dataset. The findings were that all variables possessed p-values < 0.05, leading to the null hypothesis of normality being rejected. Visual inspection was also conducted with quantile-quantile (Q-Q) plots, which enabled comparison between the sample distribution and an imaginary normal distribution. This is a vital step because most statistical methods, such as linear regression assume normality of the data. Therefore, understanding the distribution of variables directs the selection of appropriate statistical models or transformation techniques in the analysis.

#### D. Imputation Techniques for Missing Values

To ensure dataset completeness and to prevent data loss [18], a targeted imputation strategy was adopted according to variable type. For numeric missing values were imputed using column wise median imputation as it is robust to skewness and better preserve the central tendency in non-normal data. This method will maintain the overall statistical distribution of data. For categorical missing values were tagged as unknown. This strategy prevents the removal of potentially useful records. Additionally, it also enables clear accounting for data gaps

downstream models and visualisation. This dual strategy guaranteed that the dataset integrity was maintained across all merged source and no rows were removed unnecessarily, supporting reliable downstream analysis and modelling.

#### E. Feature Engineering

A comprehensive feature engineering framework was utilised to transform raw datasets into structured and meaningful features suitable for predictive modelling and analysis. Aggregated city-based features were first extracted, including the number of reported crimes by city, average tourism rating derived from reviews of individual tourist destinations, and average hotel rating derived from reviews of individual OYO hotels. Moreover, sentiment analysis of user created textual reviews was classified into four thematic categories of sentiment: bad review, good review, street scam and fraud, and drug safety and violent crimes. The number of reviews within each sentiment category was computed separately for both tourism and hotel datasets for each city, thereby enabling fine-grained analysis of public opinion and user experience across different domains. Additionally, offences were categorized into three risk levels high, medium, and low based on the seriousness and nature of the crimes. High-risk crimes included such categories as homicide, sexual assault, kidnapping, and gun crimes; medium risk crimes were such as assault, burglary, fraud, and drug crimes, while low-risk crimes were primarily traffic cases. This risk-based classification was used to categorize cities according to their dominant crime profile, which assigned each city a categorical risk rating: high, medium, or low. These constructed attributes in combination with each other present a good approximation of each city's safety, reputation, and quality of service, forming an important foundation for further analysis and modelling.

#### F. Cluster Analysis

Cluster analysis was implemented to identify natural patterns among Indian cities according to crime records, tourism satisfaction and accommodation quality. Due to the skewed distribution of several ratings and review variables log transformation was applied to reduce skewness. z-score normalisation was applied after transformation to guarantee that each numeric variable made an equal contribution to the clustering process. K-mean clustering is applied to the normalised dataset. To balance within cluster variation and interpretability, the ideal number of clusters determined through Elbow method. K-Mean clustering was then conducted with three centres, and results were visualized using Principal Component Analysis (PCA) to capture the majority of features variance. The resulting clusters revealed distinct patterns among cities with differing crime profiles and safety ratings, enabling meaningful classification for subsequent analysis.

#### G. Machine Learning and Deep Learning Techniques Used

This research applies various machine learning techniques to classify user satisfaction and crime related information from user reviews through sentiment analysis, crime trend forecasts for the upcoming years, and safety score prediction of cities. Each task uses different data preprocessing, model training, and evaluation strategies, demonstrating the flexibility of machine learning to deal with actual problems.

#### 1) BERT-Based Sentiment Analysis

To classify user generated text reviews into significant categories, the BertForSequenceClassification model from Hugging Faces Transformers library was fine tuned. This model extends the pretrained bert-base-uncased architecture by adding a classification layer to output predictions for four sentiment classes: Street Scams& Fraud, Drug Safety and Violent Crimes, Good Reviews and Bad Reviews. A maximum sequence length of 128 tokens were used to encode the reviews after tokenised using the BERT tokenizer. The model was fine-tuned from the BERT based model Bert-base-uncased by adding a classification layer to output four sentiment classes. “BERT consists of Transformer encoder, which receives the input from embedding layer and flows through the multiple self-attention layer to identify the appropriate relations between words in a text” [19]. The CrossEntropyLoss function and AdamW optimiser were used to train the model across five epochs with a batch size of 32. A training-validation ratio of 80-20 is used to the dataset. To guarantee the correct convergence, accuracy and loss measures were monitored throughout the training. The fine-tuned model is then used to classify reviews from accommodation and tourism datasets. To create a structured sentiment distribution that was incorporated into the master dataset, categorised sentiments were then combined at the city level. This contributed to the evaluation of each city's risk profiles, user satisfaction and safety.

#### 2) Crime Trend Forecasting

To model the temporal fluctuation of crimes within Indian cities, adopted a machine learning approach centred on Extreme Gradient Boosting (XGBoost). Crime forecasting has become widely recognised in the past few years as it helps authorities to address crimes analytically [20]. This approach is chosen because it is reliable when dealing with non-linear connections and can use to incorporate historical trends through engineered features. City and risk level specific annual crime figures were combined. The key features used for modelling included the prior year crime count, risk category, year and city. For smooth integration into XGBoost framework, Categorical variables are one-hot encoded. Standard metrics, such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) is used to assess the model after the model was trained using squared error loss function. Future crime trend for 2025 and 2026 were forecast recursively with the recent data available. Additionally, linear regression with lag and temporal factors and Exponential Smoothing (ETS) used to test XGBoost performance comparison. The forecasted data is then implemented in dashboard for future crime trend for all cities as inputs to downstream risk mapping

#### 3) Safety Score Predictions

To construct a quantitative city level safety score, a composite overall safety score is developed by weighted aggregation of normalized features from crime, tourism and accommodation data. Significant variables were normalized using min-max scaling. Domain knowledge weights were then assigned to all variables to represent their relative importance in perceived safety. The crime score was determined by assigning more weights to overall numbers of crime and violent crime in tourism and hotel reviews, while the tourism score and

accommodation score assigned more weight to positive ratings and penalized negative or scam ratings.

These three scores were then linearly combined with the following formula.

$$\text{Raw_Safety_Score} = 2 * \text{Tourism_Score} + 2 * \text{Accommodation_Score} - 3 * \text{Crime_Score}$$

These raw scores were then scaled to fixed range from 1.1 to 9.9. To validate this method of scoring and determine its predictive power, two machine learning models were implemented XGBoost Regressor and Deep Neural Network (DNN). The model trained to predict the overall safety score which is derived from crime, tourism and accommodation related features. Both models divided their data for training and testing at 80%-20%.

XGBoost, an ensemble learning algorithm based on gradient boosting were utilized due to its ability to handle imbalanced data and the detection of complex relationships. with 500 estimators, and learning rate parameter as 0.05, and max depth for each tree as 6 were utilized. MSE and R-squared ( $R^2$ ) metrics were employed to estimate model performance where the minimum errors and maximum explained variance were the aim.

A feedforward neural network is implemented using the Keras API with TensorFlow as the backend. The deep learning model based on the neural network with an input layer of 64 neurons, two hidden layers of 64 and 32 neurons, and an output layer that gave one predicted safety score. The model was optimized using Adam optimizer and Mean Squared Error (MSE) loss function. Normalization of the data was done by StandardScaler, which scaled the features to a mean of 0 and a standard deviation of 1. The neural network was trained for 30 epochs with batch size 64. Model performance was checked using MSE and R-squared metrics, and the training/validation loss curve of neural network. The predicted safety scores are visualized in the safety dashboard.

## II. RESULT AND DISCUSSION

### A) Exploratory Data Analysis

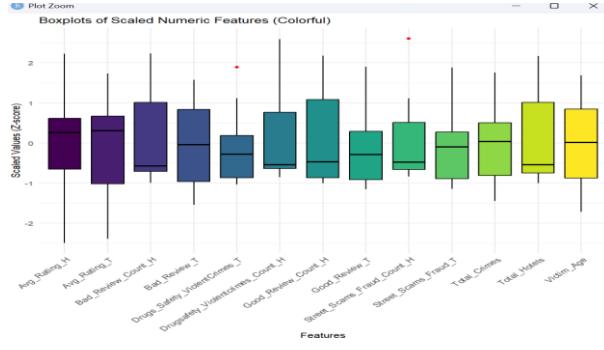


Fig. 1. Box Plot of Numerical Variables by Outcome

Boxplot of scaled numeric features reveals important characteristics of the dataset shows that spread varies across features. the variables Good Review Count Tourism, Total Hotels, and Street Scams Fraud Tourism have high variance, indicating scams, tourist service density, and attractions are very different in each city. However, the very narrow interquartile ranges of the variables such as Victim Age,

Average Rating Hotel and Average Rating Tourism indicate consistent visitor satisfaction and victim demographic activity. Outliers in some of the features such as, Good Review Count Tourism and Street Scams Fraud Tourism, represent locations that are way off the mean regarding reported street scams and visitor reviews. Observe a high degree of variation for each of the crime and safety indicators, indicating that safety issues are highly variable among cities. Such patterns, in addition to the heterogeneous urban profiles by city, provide the rationale for local safety rankings and focused policy design.

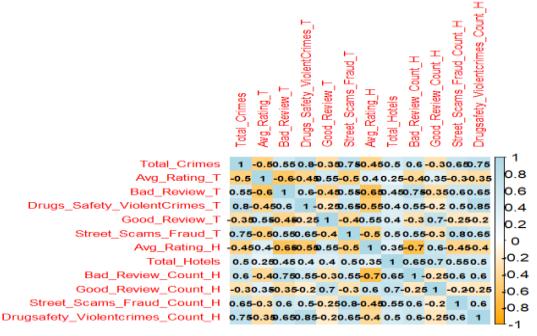


Fig.2. Correlation Heatmap

The correlation heatmap analysis provides valuable information about the relationships between crime statistics, tourist ratings, and hotel data. There was a very high positive correlation between Total Crimes and both Violent Crimes and Scam related crimes in tourism industry with respective coefficients of  $r = 0.80$  and  $r = 0.75$ , indicating that areas with high general crime levels are also heavily affected by drug crime and street crime. In addition, Bad Review by tourists had substantial positive correlations with Scam and violent crimes in tourist area ( $r=0.55$ ,  $r = 0.60$ ), indicating that adverse tourist reviews often relate to perceived or real safety concerns. Conversely, Average Tourist Ratings was negatively correlated with Total Crimes ( $r = -0.50$ ) and Bad Review ( $r = -0.60$ ), which suggests that high crime and negative reviews repress tourist overall satisfaction. These correlations are relevant to learn how crime statistics and safety perceptions impact tourism-related feedback and can be used for informing feature selection in predictive modelling.

### B. Normality Test

Normality tests indicated that most numeric variables in the data are not normally distributed. Only Victim Age, Average Tourism Rating and Average Accommodation Rating had near-normal distributions with skewness close to zero, kurtosis close to 3, and Shapiro-Wilk p-values greater than 0.05. The rest of the variables exhibited some degree of right-skewness and heavy-tailed distributions features such as Scams and Violent crimes in hotel clusters expressed through high skewness above 1.5, high kurtosis above 4, and strong Shapiro-Wilk test rejections ( $p < 0.05$ ). Visual examination through Q-Q plots and histograms verified these with substantial right-tailed departures for review count and crime related variables. These results suggest that data transformations or non-parametric methods are necessary for formal statistical analysis because parametric tests assume normality

### C. Cluster Analysis

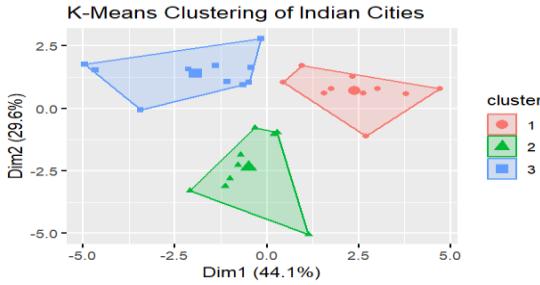


Fig. 3. K-Mean Cluster Analysis on Indian Cities

K-means Clustering is performed on standardized, and log transformed crime and review related features of Indian cities. The optimal number of clusters was determined to be three using Elbow method [21]. The first two principal components together capture 73.7% of the total variance, means the plot preserves most of the original data's structure. Cluster 1: predominantly includes cities with moderate crime rates and balanced review scores. Cluster 2: comprises cities characterised by relatively lower crime levels and better safety indicators. Cluster 3: groups cities with higher reported incidents, lower average ratings and high count of sentiment features. The identified clusters were integrated into Power BI dashboard through dynamic map visuals, allowing users to filter cities by cluster and visualize risk categories geographically using colour coded markers. Clustering offers a strong data-driven approach to identifying similar city crime and tourism profiles, which can guide urban development decision-making, police policy in relevant regions to prevent crimes [22] and tourist safety planning.

### D. Machine Learning and Deep Learning Models

#### 1.BERT-Based Sentiment Classification Performance.

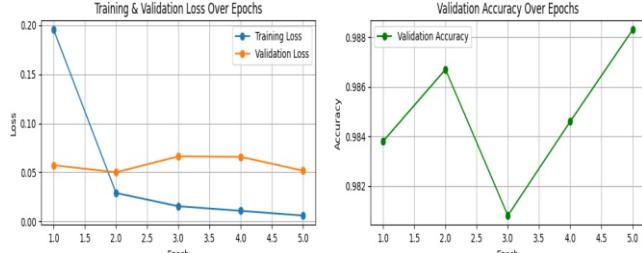


Fig. 4. Training and validation loss over epochs and validation accuracy over epochs.

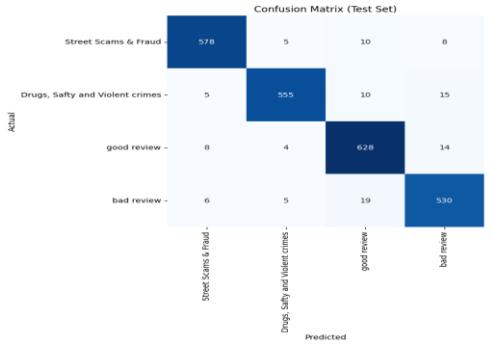


Fig. 5. Confusion Matrix

The fine-tuned BERT model achieved high performance in the multi-class classification of user reviews into four classes. The training plots and performance measures of the BERT model show its better capability in categorizing reviews. The loss plots and accuracy reflect rapid convergence with training accuracy from 95.17% (Epoch 1) to 99.79% (Epoch 5), while validation accuracy reached a plateau at 98.83%, reflecting minimal overfitting. The loss plots also support this, with sharp decrease in training loss from 0.1953 to 0.0060 and stable validation loss 0.05–0.06, supporting robust generalization. The precision-recall balance (0.9883) and F1-score highlight the stability of the model across all four classes. The corresponding confusion matrix (fig5) conforms strong predictive accuracy with minor misclassifications which is common in natural language tasks. The low rate of entropy, combined with high classification accuracy, reflects a high confidence prediction distribution, rather than random guessing. BERT outperforms lexicon-based methods [23], which limited to binary polarity and does not capture contextual or domain specific semantics through contextual embeddings with virtually perfect accuracy [24]. The validation performance consistency despite increased training accuracy highlights reliability for real-world use.

### 2. Crime Trend Forecasting

In predicting Indian crime trend across risk categories, the XGBoost model consistently achieved the highest predictive accuracy out of three forecasting techniques investigated. The model forecasted the total number of crimes in the country with a MAE of 6.05, MAPE of 29.5 percent and a RMSE of 9.14 respectively. The results significantly surpassed those obtained using traditional approaches. The ETS model reported an RMSE of 11.3 for total crimes, while linear regression obtained the least accurate results with an RMSE of 12.4 lacked the flexibility to capture temporal dynamics effectively.

Although ETS performed reasonably well in forecasting high and medium risk crimes univariate nature restricted the capacity to adjust to latest changes and geographical heterogeneity in crime patterns. The forecast produced by ETS depends on weighted average of historical data, where the weights drop exponentially according to the age of information [25]. The performance of XGBoost can be linked to the capacity to encode categorical factors and integrate lagged values as input features. These capabilities enable the model to generate reliable predictions for multiple risk categories.

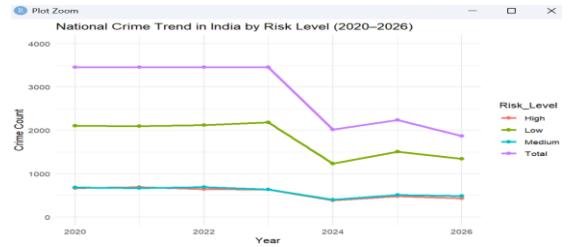


Fig.6. National Crime Trend in India by Risk Level

Figure 6 illustrates the national crime trend from 2020 to 2026. The historical and projected data are clearly aligned. The model captures a substantial decline in total and low-risk crimes

beginning in 2024 with stabilization observed towards 2026. This modelled crime trend forecasting provide meaningful insights for policymakers and reinforces the methods practical applicability in long term crime forecasting.

### 3. City Safety Score Prediction Results

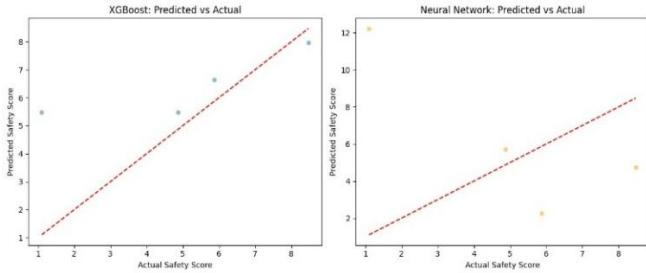


Fig.7.XGBoost and Neural Network Model Evaluation

The comparative analysis of safety score prediction models yield considerably different results for the XGBoost and Neural Network approaches. The XGBoost model demonstrated strong predictive performance with an MSE of 5.09 and  $R^2$  0.27. The result of Residual analysis shows mostly normally distributed errors centred around zero, although slight positive skewness indicating sporadic under prediction of higher values of safety scores. The Neural Network solution performed much worse in all measures. With an MSE of 26.02 and negative  $R^2$  value -2.71, suggesting that the model failed to generalise effectively. Error distribution analysis showed a heavy-tailed behaviour, with numerous large-magnitude residuals pointing towards systematic inability to predict. Training dynamics reflected unstable convergence behaviour, with validation loss reflecting large epoch-to-epoch variance. This difference in performance highlights comparative advantages of tree-based ensemble learning algorithms over deep learning algorithms to structured, medium-dimensional sparse sampling regression problems with sparse training sets. The nature feature selection qualities of XGBoost and insensitivity to noise appear particularly suited to safety score prediction, as opposed to greater parameter complexity within the neural framework against it per available training sample.

### 4.Tourism Safety Dashboard Implementation



Fig.8. Data Analysis Dashboard using Power BI

The tourism safety dashboard illustrated an integrated platform for analysing and forecasting tourism related crime patterns. Forecasted crime trend with different risk category shows the upcoming trend of crime pattern in Indian cities. Predicted city-based safety score derived from XGBoost models, identified regions with lower anticipated safety, assisting in risk prioritization. The clustering output groups cities into three distinct categories, illustrated on an interactive map for better geographical interpretation. The overall ratings and sentiment count within crime and safety in both hotels and public spaces provided users insights about hidden information often not shown by traditional tourism websites and lodging platforms. This dashboard provides users with filter options to choose city, crime category, time category, victim characteristics, and year. Through this interactivity, insights remain context-specific and aligned with the user's area of interest. Throughout different sets, the dashboard consistently indicated that crime trends vary by location and time, with identifiable patterns in age and gender distribution, hour of day, and season, especially for scams and safety concerns in both hotels and public spaces. The addition of predictive modelling gave the dashboard a useful forward-looking aspect. The models enable anticipatory decision-making on the part of tourists, authorities, and service providers. The greatest strength is adaptability through the ability of users to personalize the presentation of information to fit needs, it can act as a decision support tool capable of being suitable for multiple stakeholders. Finally, the dashboard permits safer, data-driven tourist planning by combining statistical crime data, machine learning predictions, and sentiment-driven insights into a unified, accessible interface statistical crime data.

### V. LIMITATIONS AND FUTURE ENHANCEMENTS

Despite the robust methodology and comprehensive integration, several limitations exist in the current framework. The crime data are aggregated at the city level, which hide localized crime patterns at the neighbourhood level reducing risk detail [26] due to the unavailability of granular data. Secondly, the BERT performance on sentiment classification that is applied only to English reviews restricts generalizability to multilingual situations prevalent in India. The forecasted safety score was derived using weighted equation based on domain intuition and well-informed assumptions which though reasonable but might result in bias. Despite the use of machine learning models to validate and generalise this formula, the approach may benefit from further empirical calibrations or use of ground truth safety labels in future studies. Finally, the dashboard's foundation on sporadic updates of data rather than real-time crime feeds may fail to capture the appearance of new safety threats. Addressing these limitations in future iterations could enhance the model's precision and applicability.

This research paradigm can be enriched substantially with the following strategic improvements. The dashboard accuracy and timeliness would be improved by broadening the datasets to include more sources, such as government tourism boards, international tourism reports and live feeds of police crime. More precise risk evaluation would be obtained by incorporating more geographical data such as data at the district or even

neighbourhood level would provide better crime hotspots identification. Moreover, incorporating automatic anomaly detection capabilities, predictive alert and user recommendation systems to the dashboard will dramatically enhance decision support and user interactions. To increase the access and impact the dashboard can be linked to tourism websites such as TripAdvisor and OYO or developed as a mobile application for travel planners. Expanding the NLP platform for multilingual sentiment analysis would extend coverage of diverse tourist opinions. Federated learning can be utilized to develop customized risk evaluation systems to generate customized safety advisories [27] based on user profiles such as traveller group size, gender and age while ensuring data privacy. These enhancements would individually make the system a more responsive, precise, and user-oriented safety system for tourists, and provide policymakers more subtle tools to urban safety planning and hospitality partners with actionable business insights.

## VI. CONCLUSION

This research establishes a model to assess crime and tourism connection in Indian cities, providing insights that crime patterns significantly influence tourist experiences and safety perceptions. By Combining crime statistics and tourist sentiment to identify significant spatial and temporal dynamics that influence the appeal of locations. The interactive platform translates raw safety information efficiently into decision-ready information for both travellers and decision-makers. Above all, this research establishes a methodological blueprint for future studies at the intersection of urban security and tourism economics. Clustering techniques enabled effective risk categorization of cities aiding targeted interventions. The flexibility of the framework permits the possibility to extrapolate to other geographical contexts and integrate with new technologies like real-time crime monitoring systems. The research underscores the significance of evidence-based action in tourism security management, while for the hospitality industry, it highlights the growing significance of open safety communication. The results show how machine learning, natural language processing, and visualisation tools may be used to analyse tourism safety issues. Higher geographic resolution at the level of the tourist destination, languages, and real-time data sources are possible future enhancements. Lastly, this research advances the field of city tourism resilience, providing both a practical tool for visitors and a conceptual framework for future research into smart tourism and city administration. The model provides a replicable framework for translating publicly available safety data into actionable safer data driven travel planning.

## REFERENCES

- [1] R. George, "Tourist's perceptions of safety and security while visiting Cape Town," *Tourism Management*, vol. 24, no. 5, pp. 575–585, Oct. 2003
- [2] N. Yu and J. Chen, "Design of Machine Learning Algorithm for Tourism Demand Prediction," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–9, Jun. 2022,
- [3] R. Shchokin, V. Mazychuk, O. Mykhailyk, A. Kolomiets, S. Akifzade, And Y. Tymoshenko, "The Impact Of The Crime Rate On The Hospitality And Tourism Industry In The Eu Countries," *Geojournal Of Tourism And Geosites*, Vol. 46, No. 1, Pp. 135–147, Mar. 2023
- [4] A. Lisowska, "Crime in Tourism Destinations: Research Review," *Turyzm*, vol. 27, no. 1, pp. 31–39, Jan. 2017.
- [5] E. Cohen, "Tourism-related Crime: Towards a Sociology of Crime and Tourism," *ScholarWorks@BGSU*, 2018.
- [6] R. I. Mawby and Z. I. Vakhitova, "Researching the Relationship Between Tourism, Crime and Security: The Tourism Industry and the Disenfranchised Citizens," *The Handbook of Security*, pp. 581–601, 2022.
- [7] Chhabra, Jaya; Bhattacharjee, Mahua, "Analyzing tourist preferences in India to crime and threat of terrorism," *International Journal of Research in Social Sciences*, vol. 9, no. 6, pp. 527–542, 2019.
- [8] N. Hua, B. Li, and T. Zhang, "Crime research in hospitality and tourism," *International Journal of Contemporary Hospitality Management*, vol. 32, no. 3, pp. 1299–1323, Apr. 2020
- [9] P. Brunt, R. Mawby, and Z. Hambly, "Tourist victimisation and the fear of crime on holiday," *Tourism Management*, vol. 21, no. 4, pp. 417–424, Aug. 2000.
- [10] B. Biagi and C. Detotto, "Crime as Tourism Externality," *Regional Studies*, vol. 48, no. 4, pp. 693–709, Feb. 2012.
- [11] L. Johnny and L. Jordan, "Tourism and Crime in the Caribbean: A case study of St Lucia," *Annals of Leisure Research*, vol. 10, no. 3–4, pp. 475–497, Jan. 2007.
- [12] D. P. Dash, A. K. Dash, and Yashobanta Parida, "Lethal attraction: crime as tourism externality- evidence from the Indian states," *Asia Pacific Journal of Tourism Research*, pp. 1–17, Nov. 2024.
- [13] V. Tiwari and A. Omar, "The impact of the hotel star rating system on tourists' health safety and risk perceptions: Study based on tourists' vacation experiences," *Journal of Vacation Marketing*, Jul. 2023, doi: <https://doi.org/10.1177/1356766723118880>.
- [14] Alvaro, M. Koehler, Nikolaos Konstantinou, Pavel Pankin, N. W. Paton, and R. Sakellariou, "Data Preparation: A Technological Perspective and Review," *Data Preparation: A Technological Perspective and Review*, vol. 4, no. 4, Jun. 2023.
- [15] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the Application of Deep Learning-based BERT Model in SentimentAnalysis." Available: <https://arxiv.org/pdf/2403.08217.pdf>
- [16] R. V. Kumar, "Exploratory Data Analysis using R & RStudio," Sep. 02, 2021. [https://www.researchgate.net/publication/354312419\\_Exploratory\\_Data\\_Analysis\\_using\\_R\\_RStudio](https://www.researchgate.net/publication/354312419_Exploratory_Data_Analysis_using_R_RStudio).
- [17] L. S. Nelson, "The Anderson-Darling Test for Normality," *Journal of Quality Technology*, vol. 30, no. 3, pp. 298–299, Jul. 1998, doi: <https://doi.org/10.1080/00224065.1998.11979858>.
- [18] J. M. Jerez *et al.*, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 105–115, Oct. 2010, doi: <https://doi.org/10.1016/j.artmed.2010.05.002>.

- [19] B. Selvakumar and B. Lakshmanan, "Sentimental analysis on user's reviews using BERT," *Materials Today: Proceedings*, Apr. 2022, doi: <https://doi.org/10.1016/j.matpr.2022.03.678>.
- [20] W. Safat, S. Asghar, and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 9, pp. 1–1, 2021, doi: <https://doi.org/10.1109/access.2021.3078117>.
- [21] M. Cui, "Introduction to the K-Means Clustering Algorithm Based on the Elbow Method," 2020, doi: <https://doi.org/10.23977/accaf.2020.010102>.
- [22] P. Gera and R. Vohra, "City Crime Profiling Using Cluster Analysis." Accessed: Apr. 27, 2025.
- [23] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian," *Electronics*, vol. 11, no. 3, p. 374, Jan. 2022
- [24] M. Peters, M. Neumann, L. Zettlemoyer, and W.-T. Yih, "Dissecting Contextual Word Embeddings: Architecture and Representation." Available: <https://arxiv.org/pdf/1808.08949>
- [25] G. Athanasopoulos and D. Weatherburn, "Contemporary Issues in Crime and Justice Number 219 Forecasting male and female inmate numbers: A comparison of ARIMA and ETS modelling results," 2018.
- [26] Chainey, S., & Ratcliffe, "GIS and Crime Mapping," *Google Books*, 2005 2025.
- [27] H. Lau, Y. P. Tsang, D. Nakandala, and C. K. M. Lee, "Risk quantification in cold chain management: a federated learning-enabled multi-criteria decision-making methodology," *Industrial Management & Data Systems*, vol. ahead-of-print, no. ahead-of-print, Apr. 2021