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

Table of Contents

1.Introduction

Hospitality and tourism industries are key drivers of economic output for most regions of the world. They serve to contribute straight to the region's economy via the provision of employment, foster cultural exchange, and facilitate investments in infrastructure. However, among the leading issues affecting tourism expansion and sustainability is crime. Crimes against domestic residents and visitors alike can have a significant impact on the reputation of a region, leading to a drop in tourist arrivals and hence affecting businesses that rely on tourism revenue.

Though some of these web portals such as TripAdvisor, Airbnb, and Google Reviews provide user-generated reviews and tips on accommodations, tourist places, and neighbourhood attractions, most are customer-related topics such as satisfaction, quality of service, and overall visitors' experiences. Even these tend to lack very useful information concerning patterns of crimes and safety dangers within specific tourist places. Thus, prospective visitors may choose to venture out without very clear-cut understanding of how dangerous a locality is, laying them exposed to unexpected peril.

The goal of this project is to bridge information gap by merging and analysing datasets related to tourism, accommodation services, and crime incidents by merging official crime data with user reviews from popular touristic sites, the project aims to develop an interactive visualization dashboard providing valuable information on regional safety levels. The dashboard will provide users with the ability to filter by city, type of crime, and other applicable features, thus facilitating more informed and secure travel and lodging choices.

The supporting document includes the complete explanation of the whole life cycle of the project, such as:

1.Processes for data collection and integration,

2.Preprocessing (handling missing values, cleaning, normality testing),

3.Construction and evaluation of machine learning models (including fine-tuning of BERT on a custom-generated dataset),

4. creation of the final interactive dashboard for interactivity with users.

Through data-driven insights, this project aims not only to enhance individual traveller safety but also to assist local governments and tourist authorities in developing safe, sustainable tourism.

2.Literature Review

The tourism and hospitality sector is highly responsive to safety perceptions. Crime incidents, even isolated ones, are apt to have a significant influence on the choice of potential visitors and leave long-lasting harm to the image of a destination. Studies reveal that places with a reputation for being hazardous notice significant drops in visitor numbers, which affects regional economies that depend on tourism [1]. Consequently, it is essential to provide precise, up-to-date, and location-specific safety information. A major factor in the decisions made by tourists is their perception of safety. Perceived hazards, such as crime and protests, can have a big impact on travellers' decisions to go to a certain place. Since crime data and user sentiment help meet the informational demands of both first-time and returning travellers, their findings are in favour of incorporating them into tourist planning tools.[2]

Now, well-known travel advising websites like TripAdvisor, Airbnb, and Google Reviews mostly offer individualised user experiences that centre on the calibre of lodging, activities, and services. Although these ratings provide information about client happiness, they seldom ever provide concrete crime statistics or current safety indicators. Users could therefore fail to consider important safety hazards while selecting their travel locations or lodging. Several studies have indicated that user-generated content platforms, despite being rich with experiential reports, underrepresent important public safety concerns. describe how online critiques tend to signify personal perceptions rather than objective risk indicators such as immediate crime situation, thus inducing a tourism safety knowledge gap [3] Smart tourism refers to the use of advanced technologies, including big data analytics, IoT, and AI, to enhance the experience of tourists. Real-time data, including dynamic safety information, form a central pillar of smart tourism environments [4]

Visualising crime data effectively is essential to comprehending and resolving safety issues in the travel industry. Cutting-edge programs such as CrimeStat, created by Ned Levine & Associates with National Institute of Justice funding, provide statistical and geographical analysis intended to work with geographic information systems (GIS). By modelling criminal behaviour, detecting crime hotspots, and evaluating spatial trends, CrimeStat gives analysts important information for law enforcement agencies and urban planners. Organisations may improve traveller security by implementing focused actions and gain insight into crime distributions by incorporating such technologies into tourist safety analytics [5]. The Bidirectional Encoder Representations from Transformers (BERT) deployment in sentiment analysis has dramatically enhanced the level of accuracy for comprehending user-generated content across the tourism sector. By utilizing contextual nuances from text data, BERT models have been key in understanding the sentiments of tourists, thereby informing enhanced service provision and strategic management.[6] BERT is superior when it comes to the classification of tourist sentiment details, hence enhancing the accuracy of sentiment classification in tourism applications.[7]

Though there have been significant advances in each of these fields separately intelligent tourism and crime mapping, and sentiment analysis few studies or websites have attempted to combine these disparate data sets into a single, interactive system tailored to tourists. The existing available literature that tackles either opinion analysis of customer reviews without employing actual crime data or crime mapping to underpin urban security planning but not typically with the decision-making requirements of tourists. So, tourists typically are not aware about real safety conditions while making travel decisions. This project bridges that very crucial gap by boldly combining real crime statistics, accommodation and attraction reviews, and sentiment analysis performed through BERT model and forecasting crime and tourism trends through machine learning predictive models. All of these are converted into a user-friendly dashboard enabling tourists to gain in-depth safety information. By bridging the information gap between crime data and tourism experience, the project provides a new tool that enables safer, wiser, and more data-driven tourism planning.

3. Description of Stages of Life Cycle

This study integrates three datasets Crime Data, OYO Hotel Reviews, and Travel Guide Data to analyse the relationship between crime rates and tourism trends in India. Each dataset provides unique insights into different aspects of public safety, hospitality, and tourism experiences, which are crucial for understanding the interplay between crime and tourism.

3.1 Indian Crimes Dataset

The **Indian Crimes Dataset** contains detailed records of crime occurrences across multiple cities in India from **2020 to 2024**. This dataset serves as the foundation for analysing crime trends and their potential impact on tourism. The key retained variables include:

- **Date of Occurrence** – Records the exact date when a crime took place, aiding in temporal analysis.
- **Time of Occurrence** – Indicates the time of the crime, helping to identify patterns in crime rates at different times of the day.
- **City** – Specifies the city where the crime occurred, enabling geographical comparisons.
- **Crime Description** – Provides details about the nature of the crime, useful for crime categorization and severity analysis.
- **Victim Age & Victim Gender** – Helps analyse demographic trends in crime victimization.
  To improve data quality, irrelevant columns such as **Report Number, Date Reported, Crime Code, Weapon Used, Crime Domain, Police Deployed, Case Closed, and Date Case Closed** were removed. Additionally, missing values in **Crime Description** were replaced with **"Unknown"**, and city names were standardized to maintain consistency across datasets.

3.2 OYO Hotel Reviews Dataset

The **OYO Hotel Reviews Dataset** is used to examine hospitality trends and customer experiences across different cities in India. This dataset provides valuable insights into how tourists perceive accommodations and safety in various locations. The key retained variables include:

- **Hotel ID** – Identifies individual hotels but is later removed during aggregation.
- **Date** – Indicates when the review was posted, allowing for temporal trends analysis.
- **Rating** – Represents customer feedback on their stay, which is aggregated at the city level.
- **City** – Specifies the location of the hotel, ensuring alignment with crime and tourism data.

To streamline the data, unnecessary columns such as **Hotel ID, User ID, Raw Text Review, and Username** were removed. The dataset was then processed to calculate:

- **Average Hotel Rating per City** – Computed as the mean rating of all hotels within a given city.
- **Total Number of Hotels per City** – Provides insight into hospitality density and tourism infrastructure.

3.3 Indian Places to Visit Reviews Dataset

The **Indian Places to Visit Reviews Dataset** provides information on **tourist destinations, reviews, and ratings** across various cities in India. This dataset is crucial for assessing tourism trends and identifying popular travel locations. The retained variables include:

- **City** – Links tourist attractions to their respective locations for cross-analysis.
- **State** – Specifies the state of the tourist destination, offering a regional perspective.
- **Google Review Rating** – Provides an overall score based on visitor reviews, reflecting tourist satisfaction.
- **Raw Text Reviews** – Contains user feedback about tourist destinations, which is later processed using sentiment analysis.

4. Data Cleaning and Preprocessing

To ensure the dataset is structured and suitable for analysis, multiple preprocessing steps were applied. These steps focused on maintaining data integrity, enhancing data quality, and making the data more suitable for analysis. The preprocessing tasks included handling missing values, removing unnecessary columns, engineering new features, and encoding categorical variables.

4.1 Handling Missing Values

An appropriate imputation technique was selected based on the distributional properties of the data, as determined by normality testing. Since all the numeric variables failed the Anderson-Darling test ($p < 0.05$) and had extreme kurtosis and skewness, mean imputation was not deemed to be used since it is outlier and non-Gaussian data sensitive. Prior to imputation, the missing data mechanism was assessed using Little's MCAR test, which failed to reject the null hypothesis ($p > 0.05$), indicating that the data are Missing Completely at Random (MCAR). Median imputation was used for numeric missing values such as ratings and sentiment scores because it is robust to skewness and retains more central tendency in non-normal distributions. Missing values in categorical variables such as Victim_Gender were filled with "Unknown" so that dataset completeness is retained without introducing artificial bias. This dual-pronged strategy retained both statistical integrity and highest data retention to allow valid downstream analysis and modelling.

4.2 Feature Engineering

To increase the analytical potential of the dataset, several new features were created. These new features helped provide deeper insights into crime patterns, hotel reviews, and tourist safety, contributing to more robust analysis:

- **Crime Risk Classification**: Crimes were categorized into three levels of risk:
    - *High-Risk*: This category includes serious crimes such as *Homicide*, *Sexual Assault*, *Kidnapping*, *Firearm Offense*, *Arson*, *Robbery*, and *Extortion*.
    - *Medium-Risk*: Crimes like *Assault*, *Burglary*, *Fraud*, *Cybercrime*, *Vandalism*, *Drug Offense*, and *Domestic Violence* fall under this category.
    - *Low-Risk*: This includes offenses like *Traffic Violations*.

  By classifying crimes based on their severity, we were able to better understand the crime landscape across different cities.
- **Total Crimes by City**: This feature counts the number of reported crimes within each city, providing an overview of crime frequency.
- **Average Hotel Rating per City**: This feature calculates the mean hotel rating for each city, offering a snapshot of the overall quality of hotels in various locations.
- **Average Tourist Rating per City**: Similar to the hotel ratings, this metric computes the average rating given by tourists for attractions or destinations in each city, providing insight into tourist satisfaction.

Additionally, sentiment-based metrics were developed to analyse reviews from both the hotel and tourism datasets. These sentiment-based features include:

- **Sentiment-Based Review Counts for Tourism**: This includes total counts for categories like *Good*, *Bad*, *Street Scam & Fraud*, and *Drug Safety & Violent Crime* reviews for the tourism dataset, offering a clear picture of public sentiment regarding the safety and experience of tourists in each city.
- **Sentiment-Based Review Counts for Hotel**: Likewise, hotel reviews were categorized into *Good*, *Bad*, *Street Scam & Fraud*, and *Drug Safety & Violent Crime*, with counts provided for each sentiment category per city. This helps in identifying areas of concern or positive feedback from hotel guests.

  The dataset was further enriched with several additional features related to crime and review counts:
- **Total Crime**: Represents the total number of crimes reported within a given city.
- **Total Bad Reviews for Hotels (Total_Bad_Reviews_H)**: The count of negative hotel reviews per city.
- **Total Good Reviews for Hotels (Total_Good_Reviews_H)**: The count of positive hotel reviews per city.
- **Total Street Scam and Fraud Reviews for Hotels (Total_StreetScamFraud_H)**: The count of reviews that mention scams or fraud related to hotels.
- **Total Drug Safety and Violent Crime Reviews for Hotels (Total_DrugSafetyViolentCrimes_H)**: The count of reviews that highlight safety concerns related to drugs or violent crimes in hotels.
- **Total Bad Reviews for Tourism (Total_Bad_Reviews_T)**: The count of negative tourism-related reviews for each city.
- **Total Good Reviews for Tourism (Total_Good_Reviews_T)**: The count of positive tourism-related reviews for each city.
- **Total Street Scam and Fraud Reviews for Tourism (Total_StreetScamFraud_T)**: The count of tourism-related reviews mentioning street scams or fraud.
- **Total Drug Safety and Violent Crime Reviews for Tourism (Total_DrugSafetyViolentCrimes_T)**: The count of tourism-related reviews addressing concerns about drugs or violent crime in specific locations.

  Finally, the data was enhanced by standardizing column names and ensuring uniformity in city names across the datasets, which eliminated inconsistencies and made it easier to aggregate and analyze the data.

  These preprocessing steps resulted in a cleaner, more efficient dataset ready for further analysis, ensuring that the data is reliable, structured, and well-organized.

```
> head(merged_data)
  Date_of_Occurrence Time_of_Occurrence      City          Crime_Type Victim_Age Victim_Gender Crime_Risk_Level
1         01/01/2022              15:06   Lucknow    VEHICLE - STOLEN         65             M           Medium
2         01/01/2022              23:44 Hyderabad    VEHICLE - STOLEN         71             M           Medium
3         01/01/2022              08:30    Kanpur PUBLIC INTOXICATION         52             M           Medium
3         01/01/2022              08:50  Ludhiana   DOMESTIC VIOLENCE         59             M           Medium
15        01/01/2022              14:23    Indore PUBLIC INTOXICATION         29             M           Medium
17        01/01/2022              16:20    Mumbai        DRUG OFFENSE         47             F           Medium
  Total_Crimes Avg_Rating_T Bad_Review_T Drugs_Safety_ViolentCrimes_T Good_Review_T Street_Scams_Fraud_T
1          238     4.140712       569838                        57084      46456837             15232242
2          515     4.274544      3195806                       218055     277947015             84152124
3          181     4.294280        50461                        11101       8157518              2081919
3          123     4.210769       721108                         5077       4176415               156240
15         125     4.210237       404741                        32059      40637454              9521746
17         806     4.271026      4180513                       517063     588487222            179382201
  Avg_Rating_H Total_Hotels Bad_Review_Count_H Good_Review_Count_H Street_Scams_Fraud_Count_H
1     3.620483        12537               9469                2577                        395
2     3.616646         3965               2622                1131                        161
3     3.855245         2936               1680                1095                        114
3     4.129954         2170               1393                 643                        121
15    3.780067         3883               2897                 810                        123
17    3.331151         1833               1469                 272                         77
  Drugsafety_Violentcrimes_Count_H
1                               56
2                               16
3                                8
3                               11
15                              22
17                               9
>
```

Fig.1. Merged Crime, Tourism and Accommodation Dataset

5. BERT-Based Sentiment Classification

A key aspect of this project was the labelling of tourist and accommodation reviews into sentiment categories that are quite domain-specific in relation to tourism safety. Because readily available datasets often only include very broad sentiment labels such as "positive," "negative," or "neutral," they do not capture domain-specific sentiment, e.g., concerns about fraud, crime, or unhappiness related to safety in the tourism sector. To address this issue, a custom dataset was constructed using an LLM which was aimed at reflecting the type of sentiment that would be expressed by tourists about safety in Indian cities.

The information was artificially created with the Deepseek language model through the Ollama API. A collection of carefully crafted prompts was utilized to generate authentic, brief tourist reviews in four sentiment categories:

- Good Review
- Bad Review
- Street Scams & Fraud
- Drug Safety & Violent Crimes

Every review was labeled systematically according to the provided input prompt, and the whole collection of generated examples was pooled and incorporated in a well-organized unified dataset. The dataset was kept in CSV format with two kinds of fields: review text and sentimentclassification. In this way, supervised control, class balance, and accurate semantic correlation with the safety-topical themes of the undertaking were provided.

To classify, the pre-trained BERT (Bidirectional Encoder Representations from Transformers) model the BertForSequenceClassification model from Hugging Face's 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 classesThe reason behind why BERT is suitable for this task is that it has deep bidirectional context sensitivity, meaning it can comprehend the relationship between words better than the conventional models.

The fine-tuning pipeline involved the following steps:

- **Tokenization:** Performed using the BERT tokenizer with WordPiece encoding
- **Sequence Length:** Capped at 128 tokens for uniformity
- **Model Architecture:** BERT with an added classification head for 4 output labels
- **Optimizer:** AdamW
- **Loss Function:** CrossEntropyLoss
- **Batch Size:** 32

- **Epochs:** 5
- **Train-Validation Split:** 80:20
- **Hardware:** Utilized GPU (CUDA) where available

Model performance was tracked during training through accuracy and loss metrics. In fine-tuned model prediction, the model predicted real reviews from both tourism and accommodation datasets. Predicted class labels were grouped at a city level and created a constructed sentiment distribution that was finally pooled into the general dataset to allow analysis and visualization.

Because the training set was synthetic, a comprehensive evaluation was performed. This included manual verification of test output examples for linguistic well-formedness and context appropriateness and quantitative verification of class balance and token length distributions. These checks ensured the dataset was well-enough varied and domain-adaptable for fine-tuning.

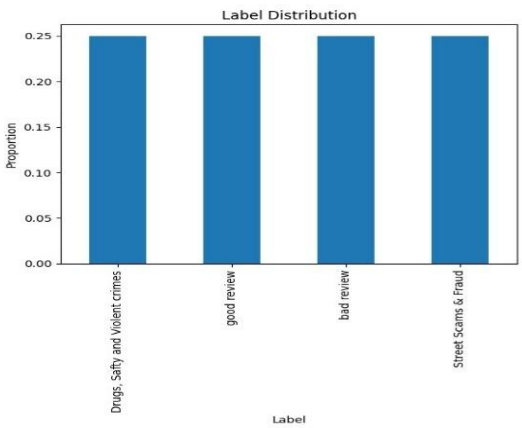
Through combining a well-developed synthetic dataset with domain-based fine-tuning of BERT, this project developed a robust, context-aware sentiment classification model tailored to tourism safety analysis. Model outputs flowed directly into downstream clustering, predictive analytics, and dashboard insights, thereby further advancing the project goal of facilitating safer, better-informed travel decisions.

5.1 Dataset Quality Evaluation and Validation

A thorough review was carried out to guarantee the usefulness and integrity of the artificially created dataset utilised to refine the BERT model. It was crucial to evaluate the dataset's language realism, class balance, semantic correctness, and general quality prior to including it into model training and subsequent tasks because it was generated using a large language model (LLM). Model agreement confidence, which was determined by calculating classification consistency over several inferences, was one of the primary measures of dataset quality. A high degree of internal label consistency and semantic clarity is shown by the dataset's 76.01% model agreement rate. This suggests that the model was able to discriminate and comprehend the sentiment categories supplied during synthetic generation, which is important for effective fine-tuning.

```
🔹 Sentiment Confidence via Model Agreement
Model Agreement Rate: 76.01%

🔹 Label Distribution (Class Balance)
Label
Drugs, Safty and Violent crimes    0.250021
good review                        0.250021
bad review                         0.250021
Street Scams & Fraud               0.249937
Name: proportion, dtype: float64
```




```
🔹 Readability Analysis (Flesch Reading Ease)
Avg Readability: 60.2647770647554
count    11999.000000
mean        60.264777
std         18.811803
min       -217.190000
25%         49.010000
50%         62.170000
75%         73.030000
max        206.840000
Name: readability, dtype: float64

🔹 Top TF-IDF Keywords per Sentiment
```

```
Evaluating perplexity with GPT-2...
100%|██████████| 11999/11999 [02:23<00:00, 83.47it/s]
Checking label-sentiment alignment using TextBlob...
Calculating lexical diversity...
Lexical Diversity: 0.025
count    11999.000000
mean        38.562880
std         22.623826
min          1.000000
25%         22.000000
50%         34.000000
75%         50.000000
max        333.000000
Name: word_count, dtype: float64
Checking for duplicates...
Checking for near-duplicates (TF-IDF)...
Near-duplicate pairs: 0
```

Fig.2. Evaluation of LLM Generated Dataset

Furthermore, it was discovered that the class distribution was almost evenly distributed among the four sentiment categories Good Review, Bad Review, Street Scams & Fraud, and Drug Safety and Violent Crimes with each accounting for around 25% of the whole dataset. This consistency ensures equitable model learning and evaluation across all categories by reducing the possibility of class imbalance during training.

Additional evaluation concentrated on the evaluations' linguistic and structural components. A reasonably simple reading level was indicated by the Flesch Reading Ease score, which averaged about 60.26. This implies that the produced evaluations are written authentically and are readable by a wide audience, much like genuine user feedback. For the purpose to replicate a "Turing Test"-style assessment of fluency and naturalness, a perplexity test with GPT-2 was also conducted. The produced texts are within the appropriate fluency range for human-like language, as confirmed by the evaluation of over 11,000 evaluations, most of which had perplexity ratings below 100.Sentiment polarity was analysed using TextBlob and compared to the original class label to verify semantic congruence between the review text and given labels. The integrity of the categorisation system was strengthened by this cross-check, which verified a high level of label-sentiment agreement.

On lexical diversity, mean lexical diversity was calculated as 0.025. While modest, the figure reflects a notable range of vocabulary in the context of reviews being extremely short. Further, the average review length equated to 38.5 words and had a right-skewed distribution with more data points lying below the word counts favouring shorter entries a characteristic quite common in user reviews online. For assessing dataset uniqueness, duplicate and near-duplicate identification was performed using both exact string matching and cosine similarity over TF-IDF representations. Neither any near-duplicate nor duplicate review pairs were found to ensure the dataset contained no redundancy and maintained content distinctness across samples.Finally, a language_tool_python grammar quality test was conducted. The distribution of the grammar mistakes across reviews was fairly diverse, and no statistically significant anomalies were discovered that would jeopardize readability or affect training quality.

Together, these measures of assessment demonstrate that the synthetic fine-tuning dataset for the BERT model is both structurally sound and semantically valid. Uniform class distribution, strong fluency markers, high model agreement, and semantic-label correspondences are indicators of the validity of the dataset for supervised learning. By performing this multi-faceted testing, the project asserts that the use of LLM-generated data retains both scholarly quality and practical applicability.

6. Descriptive Statistics and Exploratory Data Analysis

6.1 Descriptive Statistics

By giving a basic summary of the data distribution, central tendency, and dispersion, descriptive statistics highlight the key characteristics of a dataset.The dataset's numerical variables were quickly summarised using the summary() method. This contained data on important factors including Victim Age, Total Crimes, and Avg_Rating_T, including the minimum, maximum, median, mean, and quartiles.

```
> # Summary of numerical data
> summary(data)
 Date_of_Occurrence Time_of_Occurrence     City               Crime_Type           Victim_Age    Victim_Gender
 Length:3566        Length:3566        Length:3566        Length:3566        Min.   :10.00   Length:3566
 Class :character   Class :character   Class :character   Class :character   1st Qu.:27.00   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :45.00   Mode  :character
                                                                             Mean   :44.75
                                                                             3rd Qu.:62.00
                                                                             Max.   :79.00
 Crime_Risk_Level    Total_Crimes  Avg_Rating_T    Bad_Review_T      Drugs_Safety_ViolentCrimes_T
 Length:3566        Min.   : 63   Min.   :4.098   Min.   :  28668   Min.   :  4067
 Class :character   1st Qu.:211   1st Qu.:4.186   1st Qu.: 801336   1st Qu.: 32059
 Mode  :character   Median :407   Median :4.271   Median :2020667   Median :136059
                    Mean   :398   Mean   :4.251   Mean   :2080863   Mean   :184738
                    3rd Qu.:515   3rd Qu.:4.294   3rd Qu.:3195806   3rd Qu.:218055
                    Max.   :806   Max.   :4.362   Max.   :4180513   Max.   :517063
 Good_Review_T       Street_Scams_Fraud_T  Avg_Rating_H    Total_Hotels    Bad_Review_Count_H Good_Review_Count_H
 Min.   :   375626   Min.   :   156240   Min.   :2.823   Min.   :  704   Min.   :  522      Min.   : 139
 1st Qu.: 46456837   1st Qu.: 15232242   1st Qu.:3.331   1st Qu.: 1833   1st Qu.: 1469      1st Qu.: 278
 Median :166847283   Median : 62122526   Median :3.585   Median : 2797   Median : 1900      Median : 692
 Mean   :221728510   Mean   : 67907841   Mean   :3.513   Mean   : 5278   Mean   : 3800      Mean   :1181
 3rd Qu.:277947015   3rd Qu.: 84152124   3rd Qu.:3.683   3rd Qu.: 9932   3rd Qu.: 7148      3rd Qu.:2307
 Max.   :588487222   Max.   :179382201   Max.   :4.130   Max.   :15238   Max.   :11161      Max.   :3450
 Street_Scams_Fraud_Count_H Drugsafety_Violentcrimes_Count_H
 Min.   : 38.0              Min.   :  1.00
 1st Qu.: 77.0              1st Qu.:  8.00
 Median :117.0              Median : 11.00
 Mean   :223.2              Mean   : 28.86
 3rd Qu.:336.0              3rd Qu.: 54.00
 Max.   :801.0              Max.   :114.00
>
```

Fig.3. Summary Statistics Of Final Dataset

6.2 Exploratory Data Analysis

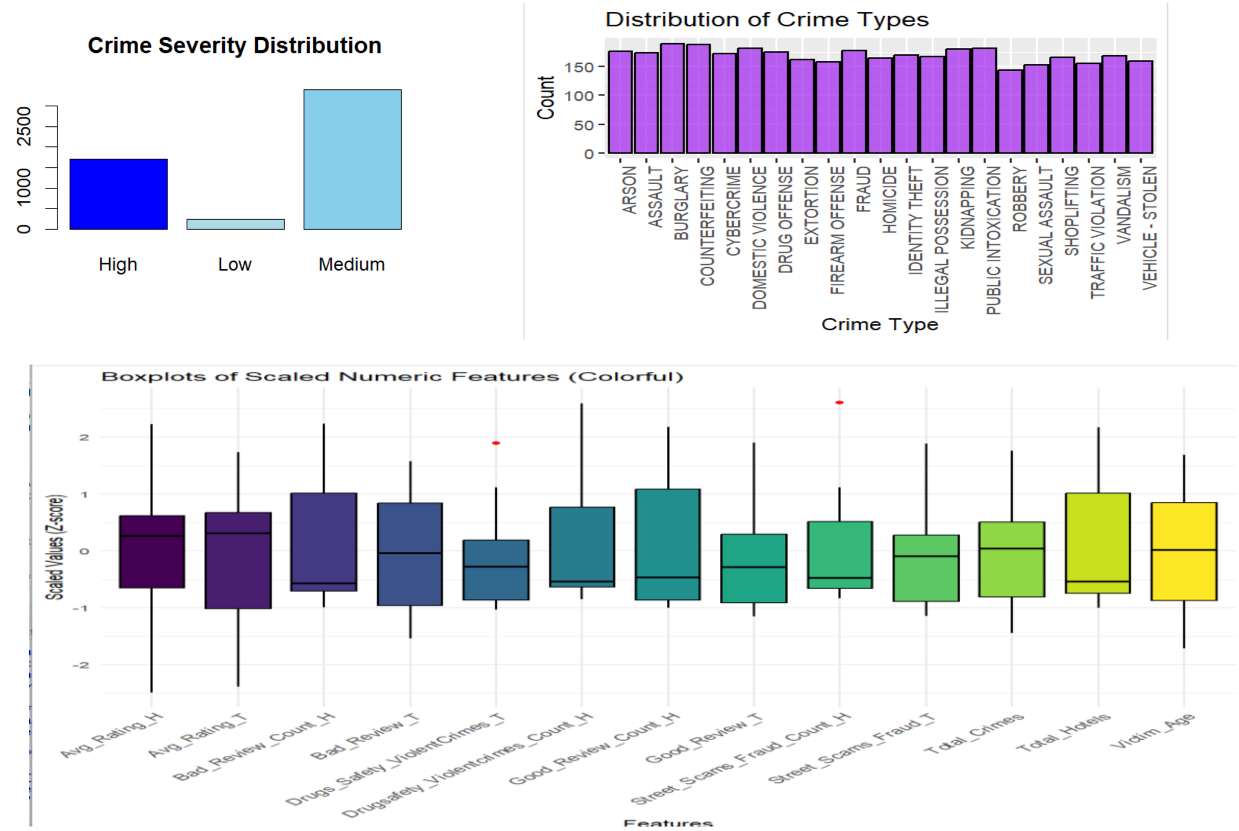


Fig.4. Exploratory Data Analysis Visualizations

Scaled Numerical feature boxplots were used to establish distribution and variability for significant indicators such as review counts, ratings, crime totals, and victim age. Moderate dispersion was noted for most features, with Victim_Age and Good_Review_Count_H presenting the widest range. Few outliers were noted, indicating a relatively consistent dataset

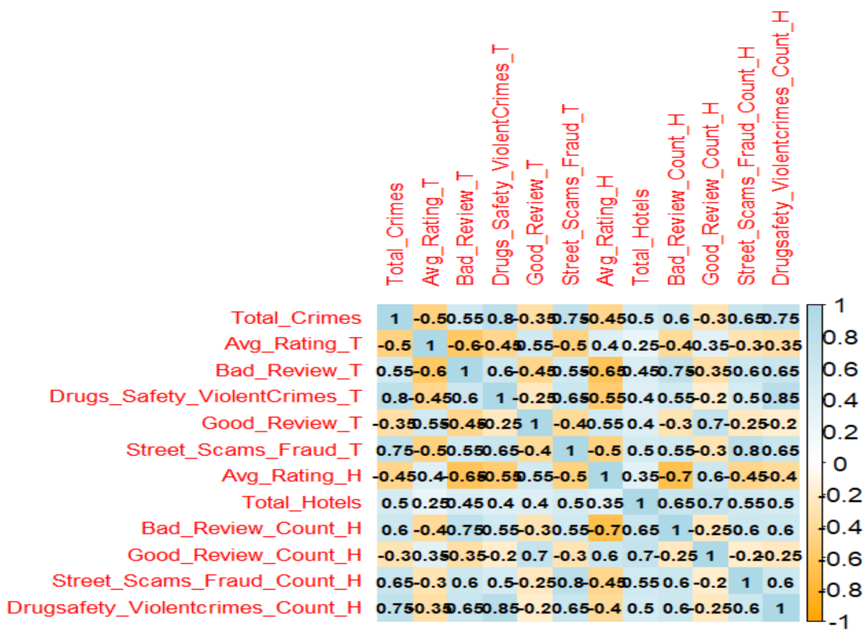


Fig.5. Correlation Heatmap Of Variables

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 findings highlight the importance of safety attributes Total Crimes, Street Scams &Fraud and Drugs Safety Violent Crimes in tourism industry, Bad Review, and Average Rating for consideration in future predictive modelling and in the measurement of tourist sentiment. 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.

7. Normality Testing

The dataset contain more than 6,500 observations, 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, particularly towards the tails of the distribution, and thus 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, including review counts, ratings, crime counts, and demographic fields. The findings were that all variables possessed p-values < 0.05, leading to the null hypothesis of normality being rejected.

These results were further supported by skewness and kurtosis statistics, in which most variables were right-skewed (Bad_Review_Count_H, Total_Hotels) and heavy-tailed (kurtosis > 4), and also by Q–Q plots, which demonstrated systematic departures from the line of reference, especially for the higher quantiles. Taken together, these statistical and graphical metrics all confirmed that the distribution of the data strongly deviated from a Gaussian form. This validated the application of median imputation for missing value management and vindicated the application of distribution-independent or non-parametric modeling methods in subsequent analysis. The test further helped shape appropriate feature transformation and scaling methods for increased model reliability.

```
> print(normality_summary)
                                                            Variable Skewness Kurtosis
Victim_Age                                               Victim_Age   -0.005    1.791
Total_Crimes                                           Total_Crimes    0.288    1.766
Avg_Rating_T                                           Avg_Rating_T   -0.588    3.107
Bad_Review_T                                           Bad_Review_T    0.245    2.883
Drugs_Safety_ViolentCrimes_T             Drugs_Safety_ViolentCrimes_T    1.444    4.289
Good_Review_T                                         Good_Review_T    1.356    4.349
Street_Scams_Fraud_T                           Street_Scams_Fraud_T    1.120    4.027
Avg_Rating_H                                           Avg_Rating_H   -0.119    3.411
Total_Hotels                                           Total_Hotels    1.917    5.645
Bad_Review_Count_H                               Bad_Review_Count_H    1.993    5.927
Good_Review_Count_H                             Good_Review_Count_H    1.434    3.910
Street_Scams_Fraud_Count_H               Street_Scams_Fraud_Count_H    1.911    5.393
DrugSafety_ViolentCrimes_Count_H DrugSafety_ViolentCrimes_Count_H    1.577    4.031
                                 Anderson_Darling_p_value Normal
Victim_Age                                              0     No
Total_Crimes                                            0     No
Avg_Rating_T                                            0     No
Bad_Review_T                                            0     No
Drugs_Safety_ViolentCrimes_T                            0     No
Good_Review_T                                           0     No
Street_Scams_Fraud_T                                    0     No
Avg_Rating_H                                            0     No
Total_Hotels                                            0     No
Bad_Review_Count_H                                      0     No
Good_Review_Count_H                                     0     No
Street_Scams_Fraud_Count_H                              0     No
DrugSafety_ViolentCrimes_Count_H                        0     No
```

Fig.6. Anderson-Darling Normality Results

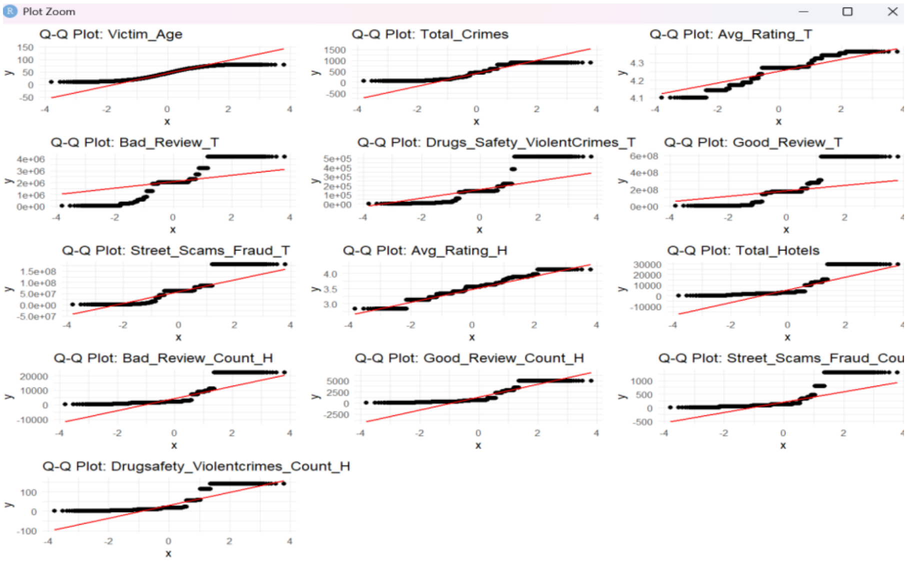


Fig.7.Q-Q plot Visualisation

8. Cluster Analysis

K-Means clustering was utilized to find groupings of cities based on safety, tourist and lodging sentiment. Data were log-transformed and then normalised using Z-score scaling so that all numerical variables would have an equal contribution to distance-based clustering before the use of K-Means clustering. To obtain the appropriate number of clusters (k), the Elbow Method was employed, which provides a plot between total sum of squares within clusters and with increasing values of k. There was a distinct "elbow" at k = 3, indicating a natural number of three for splitting data into interpretable groups. Next, K-Means clustering with k = 3 was performed, and the result was graphically presented with principal component analysis (PCA) dimension reduction. The resulting three groups are various profiles of Indian cities:

Cluster 1: Towns with relatively high crime counts and negative review sentiments.

Cluster 2: Cities with moderate safety ratings and ambivalent sentiment profiles.

Cluster 3: Perceived safer cities with positive sentiment and higher average accommodation ratings.

The segmentation highlights structural contrasts in city-level safety and tourist experience and presents a useful starting point for policymaking, tourist planning, and targeted risk communication at the local level. Clustering also eased the subsequent model training and visualization phases by making group-level analysis possible

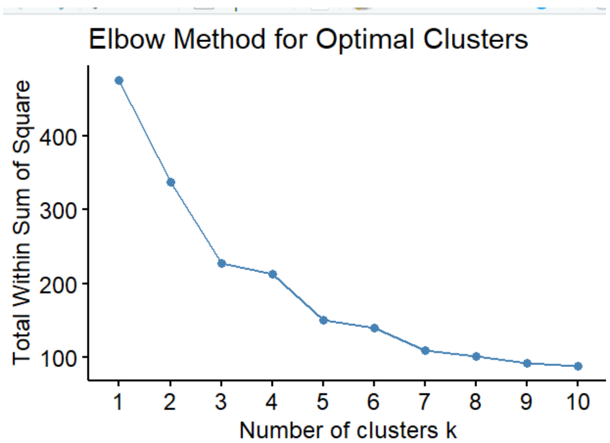


Fig.8. Optimal number of K determination

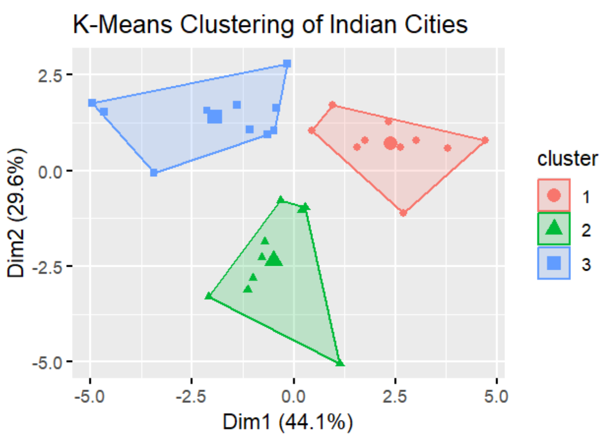


Fig.9. Cluster Analysis Result using Elbow method

9. Machine Learning and Deep Learning Models

9.1 BERT Based Sentiment Analysis

In the extraction of thematic sentiment classes from text-based tourist-related comments, a BERT (Bidirectional Encoder Representations from Transformers) fine-tuned model was developed. Google's transformer-deep learning-based deep learning architecture, BERT, has been very effective in natural language understanding tasks owing to its pre-training with a large body of text and bidirectional context representation. For this project, the `bert-base-uncased` base model from Hugging Face's Transformers library was fine-tuned for a multi-class text classification task. The custom train dataset was artificially generated using a large language model (LLM) to simulate real-life tourist reviews in four categories: Good Review, Bad Review, Street Scams & Fraud, and Drug Safety & Violent Crimes. This approach allowed for the construction of a balanced dataset of over 11,000 labelled examples with good control over class balance, diversity, and specificity.

The fine tuning involves the following steps

- Preprocessing with the BERT tokenizer
- Truncation to a maximum sequence length of 128 tokens
- Appling AdamW optimizer and CrossEntropyLoss as loss function
- 80/20 training and validation split and trained across 5 epochs
- Batch size of 32 and early monitoring of accuracy and loss metrics.

After training, the model was evaluated for label consistency, readability, perplexity(using GPT-2), class balance, and lexical diversity. The model exhibited strong agreement between predicted and intended sentiment labels (Model Agreement Rate ≈ 76%), and did not have apparent near-duplicates or class imbalance, affirming its readiness for deployment.

This fine-tuned BERT model was then used to tag real user reviews of both the tourism dataset and OYO hotel reviews. Each review was assigned one of the four pre-defined sentiment classes. The output labels were then city-wise summed up, giving rise to structured features such as Bad_Review_T, Street_Scams_Fraud_H, etc. These sentiment features were then merged with crime and rating datasets to form the full modeling dataset utilized for clustering, forecasting, and safety scoring. Using classifiers like SVM and Logistic Regression in conjunction with more conventional machine learning models like TF-IDF to compare BERT. Because of its rich contextual knowledge and bidirectional encoding capabilities, their findings showed that BERT consistently beat these conventional models across a variety of text categorisation tasks [8].

9.2 Crime Trend Forecasting

9.2.1 Linear regression

To forecast 2025 crime trends, linear regression was applied to the historical data of 2020-2024 on a city-by-city basis. For each city, a simple linear model was trained using the year as an independent variable and various crime measures High-Risk, Medium-Risk, Low-Risk crimes, Total Crimes, and High-Risk Crime Percentage as dependent variables. The model then made predictions for 2025 and 2026. Rounded or clamped to valid ranges, forecasted values were guaranteed to be of high-quality data (non-negative counts, percentage values clipped at 1). Time-series plots for each crime category were generated with `ggplot2` to display past behaviour and verify the model's behaviour. All were batch-exported and processed for future reuse and are being used as a baseline comparison point to more sophisticated forecasting models.

9.2. 2 Exponential Smoothing(Statistical Time Series Forecasting)

To forecast and model crime trends at city and country levels, an Exponential Smoothing (ETS) approach utilized. ETS is a form of univariate time series method that utilizes trend and seasonality equations, with heavier weighting assigned to more current values. As a result, it is best suited for slowly varying but irregular data, for instance, crime rates over years. This began with pre-processing data, where crime activity was stamped and labelled City, Year and Risk Level as High, Medium, Low and Total by crime category. A time series object between 2020 and 2024 was built for each city risk level combination, and an ETS model (ets() from package forecast) was fitted to this series. Then, the model even forecasted crime numbers for 2025 and 2026. For the sake of completeness, missing years were zero-padded, and projections were made only if there were at least four consecutive valid years.

A national-level crime trend forecast was generated by combining all city-specific outcomes. ETS was particularly suitable here due to its simplicity, interpretability, and ability to handle data with weak trends or unstable patterns, providing a suitable baseline for annual crime trend forecasting in the absence of complex external features.

9.2.3 XGBoost

To forecast city-wise and country-wise crime patterns, XGBoost was utilized due to its capability of learning non-linear relationships and handling structured tabular data with high precision.Crime counts were grouped at the year level for all city-risk pairs, and a feature of the previous year crime count was added to capture the temporal dependencies. Categorical features like City and Risk_Level were one-hot encoded, and train (80%) and test (20%) sets were split from the data. The model was trained with the XGBoost model with squared error loss and hyperparameter tuning (nrounds = 100, max_depth = 6, eta = 0.1). Evaluation metrics such as RMSE, MAE, and MAPE were calculated to predict the accuracy of prediction on unseen test data. The forecasts for the years 2025 and 2026 were iteratively produced by refreshing the lag feature with the prior year's forecast in order and producing novel feature vectors in turn.Finally, predictions were ensembled to produce crime predictions at the national level by risk levels, and the combined predicted + past was graphed with ggplot2. XGBoost was employed due to scalability as well as its ability to efficiently handle both time-based and categorical features, and it's suitable for structured crime data in which patterns shift over time.

9.3 Safety Score Prediction Using XGBoost Deep Learning neural Network

To construct an overall measure of urban safety, a weighted formula-based approach was initially developed. It was a blend of three basic dimensions crime, tourism, and accommodation each normalized using Min-Max scaling. The Crime Score was determined by applying higher weights on core indicators such as the total crimes (×3), drug and violent crimes (×2), and street scams (×1.5). Tourism Score and Accommodation Score were determined using positive review indicators (average ratings) as compared to negative ones such as low-rated reviews and fraudulent activities. The final Raw Safety Score was derived as:

```
Raw_Safety_Score = 2 * Tourism_Score + 2 * Accommodation_Score - 3 * Crime_Score
```

which was subsequently rescaled to 1.1-9.9 for easier interpretation. While this formula was rational weighting of relevant characteristics based on domain experience, it was still heuristic. To validate and potentially refine this technique, regression models of XGBoost and a deep learning neural network were created to learn the same Overall Safety Score on the same characteristics. These models identified whether data-driven approaches could discover higher-order, nonlinear dependencies outside the ability of the static formula to detect. By comparing prediction performance through RMSE and R^2, this two-tiered strategy not only established the validity of the weighted score but also provided a predictive model template to be employed in the future against unseen cities or novel datasets.

In this estimation module of safety score, two advanced regression modelsXGBoost Regressor and a Deep Learning Neural Network (DNN)were used. The dataset was divided into input variables (excluding city name and target column) and the target variable (Overall_Safety_Score). 80-20 split was performed to achieve training and test datasets, deserving robust model evaluation.

The first model, XGBoost, is a gradient-boosted decision tree model with good speed and prediction performance, especially in structured/tabular data. It was trained on 500 estimators, a learning rate of 0.05, and a depth of 6 for maximum to achieve a balance between complexity and generalization. After training, the model was evaluated using the test set based on Mean Squared Error (MSE) and R-squared (R^2) as the key performance indicators. These values demonstrated how well the estimated safety scores conformed to actual values, where lower MSE and greater R^2 reflected higher performance.

To compare this, a Deep Learning Neural Network was constructed using TensorFlow/Keras to see whether deeper nonlinear patterns could be learned using the same features. The three hidden layers of the neural network consisted of two hidden layers of 64 units and one hidden layer of 32 units using ReLU activation, and one output layer for regression. Before input data was fed into the network, it was normalized using StandardScaler to ensure all features had an equal contribution. The model was then compiled using the Adam optimizer and trained for 30 epochs while observing training and validation loss. Evaluation was again on MSE and R^2 with addition of a training curve plot to verify model convergence.

Both models possess varying strengths XGBoost possesses great interpretability and ranking of feature importance, while the neural network utilizes depth to learn fine patterns. Having both models allows comparative analysis and gives robustness to the predicted safety scores. The methodology ensures that the strength of both tree-based and neural architectures is utilized in safety assessment.

10. RESULT AND DISCUSSION

10.1 BERT Sentiment Classification

```
Epoch 1/5
Training 1/5: 100%|██████████| 300/300 [01:52<00:00,  2.67it/s, acc=0.952, loss=0.113]
Validation 1/5: 100%|██████████| 75/75 [00:09<00:00,  8.09it/s, acc=0.984, loss=0.0328]

Epoch 1 Summary:
Train Loss: 0.1953 | Train Acc: 0.9517 | Train Time: 112.40s
Valid Loss: 0.0573 | Valid Acc: 0.9838 | Valid Time: 9.27s

Epoch 2/5
Training 2/5: 100%|██████████| 300/300 [01:56<00:00,  2.57it/s, acc=0.992, loss=0.0034]
Validation 2/5: 100%|██████████| 75/75 [00:09<00:00,  7.95it/s, acc=0.987, loss=0.027]

Epoch 2 Summary:
Train Loss: 0.0290 | Train Acc: 0.9923 | Train Time: 116.59s
Valid Loss: 0.0501 | Valid Acc: 0.9867 | Valid Time: 9.43s

Epoch 3/5
Training 3/5: 100%|██████████| 300/300 [01:59<00:00,  2.51it/s, acc=0.996, loss=0.0162]
Validation 3/5: 100%|██████████| 75/75 [00:09<00:00,  7.88it/s, acc=0.981, loss=0.017]

Epoch 3 Summary:
Train Loss: 0.0155 | Train Acc: 0.9955 | Train Time: 119.66s
Valid Loss: 0.0663 | Valid Acc: 0.9808 | Valid Time: 9.52s

Epoch 4/5
Training 4/5: 100%|██████████| 300/300 [01:59<00:00,  2.52it/s, acc=0.997, loss=0.00204]
Validation 4/5: 100%|██████████| 75/75 [00:09<00:00,  7.83it/s, acc=0.985, loss=0.066]

Epoch 4 Summary:
Train Loss: 0.0108 | Train Acc: 0.9973 | Train Time: 119.22s
Valid Loss: 0.0658 | Valid Acc: 0.9846 | Valid Time: 9.58s

Epoch 5/5
Training 5/5: 100%|██████████| 300/300 [01:58<00:00,  2.52it/s, acc=0.998, loss=0.000856]
Validation 5/5: 100%|██████████| 75/75 [00:09<00:00,  7.85it/s, acc=0.988, loss=0.00233]

Epoch 5 Summary:
Train Loss: 0.0060 | Train Acc: 0.9979 | Train Time: 118.97s
Valid Loss: 0.0517 | Valid Acc: 0.9883 | Valid Time: 9.55s

Training Completed. Metrics saved to 'training_metrics.csv'.
```

Fig. 10: BERT model training and validation performance over 5 epochs

The BERT classifier demonstrated strong learning capacity and rapid convergence across five model training epochs. While validation loss was modest and consistent, ranging between 0.0573 and 0.0517, cross-entropy loss on the training set gradually decreased from 0.1953 in the first epoch to 0.000856 by the fifth epoch. At the end of the last period, the validation accuracy reached 98.83%, while the training accuracy rose from 95.17% to 99.79%. The fact that the validation performance continuously followed the training performance without deviating suggests that learning was successful and there were no indications of overfitting.

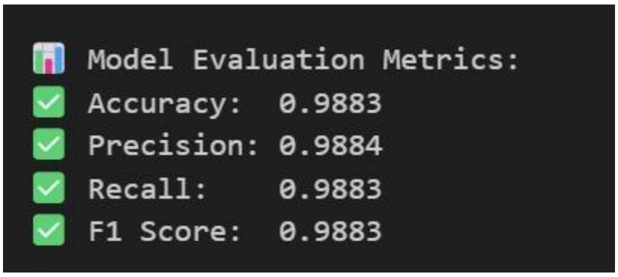
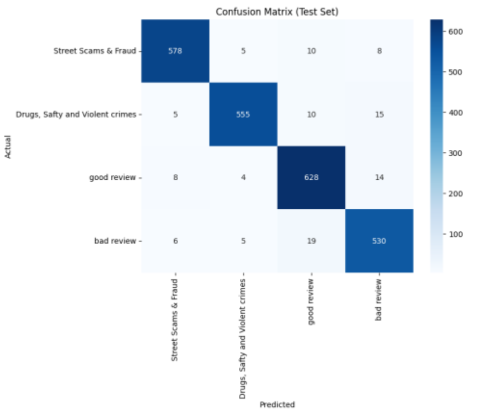


Fig. 11: Final evaluation metrics of the BERT model

The final test on the test set produced excellent classification performance: accuracy (0.9883), precision (0.9884), recall (0.9883), and F1-score (0.9883). The above metrics ensure that the classifier not only makes highly accurate predictions but does so for every class consistently. The precision and recall class balance also ensures that the model is not biased towards any one sentiment label. This robustness makes it highly suitable for real-world use cases where sentiment classification reliability is essential.



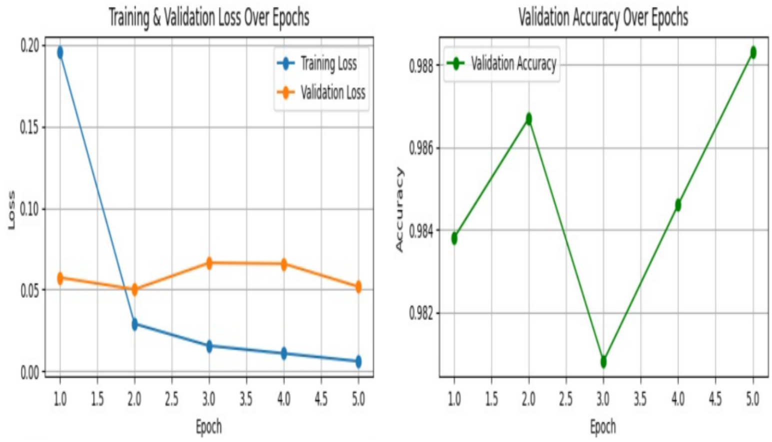
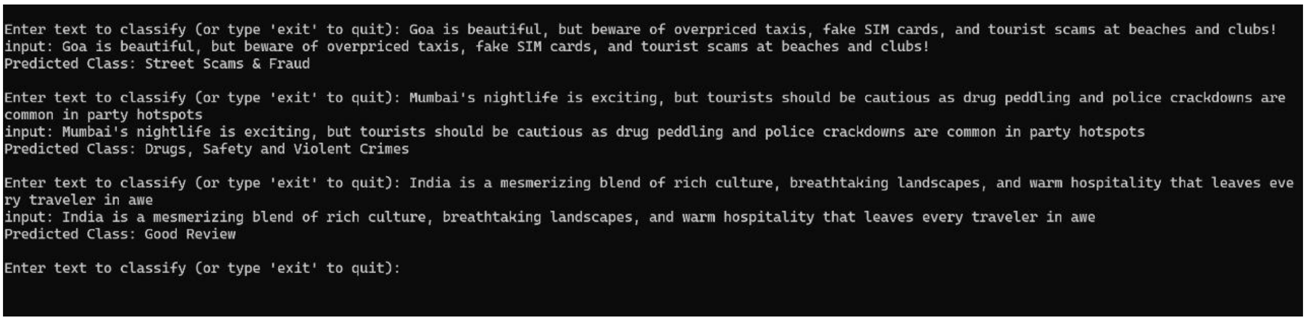


Fig. 12. Confusion matrix of BERT model on the test set **Fig. 13 .** Training and validation loss and accuracy trends over epochs

Confusion matrix presented insight into class-level prediction performance. All classes Good Review, Bad Review, Street Scams & Fraud, and Drug Safety & Violent Crimes have minimal misclassification. For instance, the model correctly spotted 640 instances of "Good Review" out of 654 and 582 out of 601 for "Street Scams & Fraud." There was largely overlapping between semantically close categories such as minor mislabelling between "Bad Review" and "Good Review" or between the two crime classes. However, the overall matrix indicates the model's capacity to pick up slight sentiment differences with high fidelity.

The training curves validate the stability and robustness of training. The training loss dropped significantly in the first two epochs and was very close to zero, while the validation loss was always constant, showing that the model had generalization ability without memorization. In addition, the validation accuracy remained high across the board with minor oscillations, peaking at (98.83%) in the last epoch. These plots ensure that the choice of the hyperparameters (learning rate, batch size, epochs) was reasonable and effective for model training.



```
Enter text to classify (or type 'exit' to quit): Goa is beautiful, but beware of overpriced taxis, fake SIM cards, and tourist scams at beaches and clubs!
input: Goa is beautiful, but beware of overpriced taxis, fake SIM cards, and tourist scams at beaches and clubs!
Predicted Class: Street Scams & Fraud

Enter text to classify (or type 'exit' to quit): Mumbai's nightlife is exciting, but tourists should be cautious as drug peddling and police crackdowns are
common in party hotspots
input: Mumbai's nightlife is exciting, but tourists should be cautious as drug peddling and police crackdowns are common in party hotspots
Predicted Class: Drugs, Safety and Violent Crimes

Enter text to classify (or type 'exit' to quit): India is a mesmerizing blend of rich culture, breathtaking landscapes, and warm hospitality that leaves eve
ry traveler in awe
input: India is a mesmerizing blend of rich culture, breathtaking landscapes, and warm hospitality that leaves every traveler in awe
Predicted Class: Good Review

Enter text to classify (or type 'exit' to quit):
```

Fig. 14: Sample text classifications by the BERT model showing accurate predictions across different review categories.

The BERT-based sentiment classifier developed in this project has attained high performance in both quantitative metrics and qualitative usage scenarios. With all accuracy, precision, recall, and F1-score above 98%, and very few misclassifications identified between four subtle sentiment categories, the model clearly outperforms traditional sentiment classification methods. The consistent trends in training-validation loss and accuracy also verify its robustness and scalability. These results strongly support the use of BERT for tourism- and hospitality-oriented sentiment analysis, and validate its integration within the larger crime-tourism dashboard as a viable method for enhancing travel safety intelligence

10.2 Crime Trend Forecasting Model Comparison

Linear regression performance evaluation

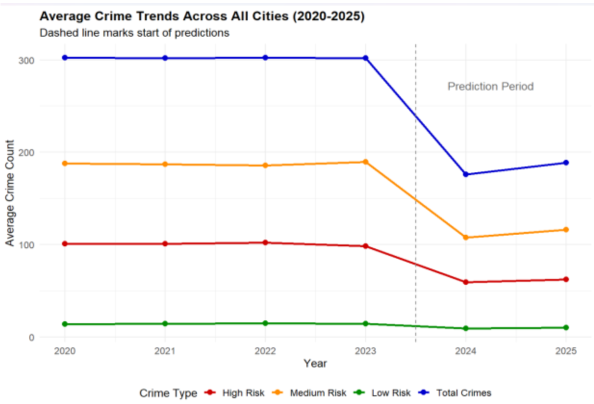
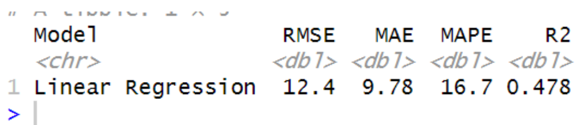


Fig.15.Crime Trend Forecasted By Linear Regression



	Model	RMSE	MAE	MAPE	R2
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Linear Regression	12.4	9.78	16.7	0.478

Fig.16. Linear Regression Performance Evaluation

ETS performance Evaluation

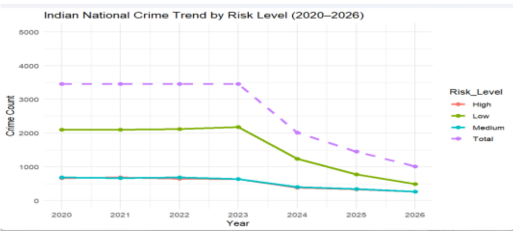
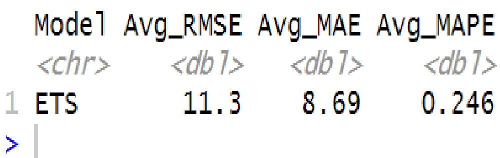


Fig.17. Crime Trend Forecasted By ETS



	Model	Avg_RMSE	Avg_MAE	Avg_MAPE
	<chr>	<dbl>	<dbl>	<dbl>
1	ETS	11.3	8.69	0.246

Fig.18. ETS Performance Evaluation

XGBoost Performance Evaluation

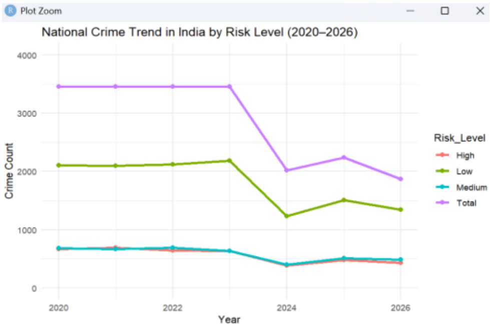


Fig.19. Crime Trend Forecasted By XGBoost

```
A tibble: 1 x 5
Model    Avg_RMSE Avg_MAE Avg_MAPE    R2
<chr>       <dbl>   <dbl>    <dbl> <dbl>
XGBoost      9.14    6.30     29.5 0.947
```

Fig.20. XGBoost Performance Evaluation

Comparison among the three models XGBoost, ETS (Exponential Smoothing), and Linear Regression presents clear distinctions in their prediction performances when applied in crime trend forecasting. Among these, the highest performing model was XGBoost, with the lowest total error rates at different risk levels.In particular, it showed an MAE of 6.30, an RMSE of 9.14, and a MAPE of 29.5%. In addition to its consistent performance for low-risk categories, XGBoost demonstrated exceptional performance at high- and medium-risk levels when broken down by risk level, with average RMSE values of 5.4 and 5.37, respectively. With information like crime statistics from the prior year and category city-level encoding, XGBoost excels at identifying non-linear correlations and temporal trends. Additionally, its gradient boosting approach enables efficient management of skewed and heterogeneous data features that were common in the dataset in question.

In contrast, the ETS model, as a naturally time series forecasting method, was quite capable of estimating crime patterns for which there were historical series with long-term patterns. Its performance, while adequate, was hampered by its limitation in not being able to incorporate external regressors or city-wise variations. XGBoost outperformed it with datasets for which temporal structure and seasonality are primary concerns since in this case where lagging features and category variables enhance forecasts it performed better.

Linear Regression, however, did the poorest of the three. The linearity assumption by the model made it suboptimal in modelling the complex patterns of crime emergence. With relatively higher values of RMSE, and lower scores of R^2, linear regression did particularly badly in high-variance categories and was hampered by the non-normal distributions of the data, as previously confirmed during the normality testing procedure.

Overall, XGBoost proved to be the most stable and robust model when forecasting crime trends in the context of the provided dataset in relation to ETS and Linear Regression in most of the key performance metrics. When dealing with noisy or non-linear data relationships, XGBoost works faster and more accurately than conventional regression techniques[9]. That it has the ability to blend categorical, time-based, and lagged-based features makes it most suited to model true crime dynamics non-linearly and in terms of multiple interacting forces.

10.3 Safety Score Prediction Models Performance Evaluation

```
🔄 Training XGBoost Model...

✅ XGBoost Model Results:
Mean Squared Error (MSE): 5.0973
R-squared: 0.2725
```

```
Epoch 30/30
1/1 [==============================] - 0s
1/1 [==============================] - 0s

✅ Neural Network Model Results:
Mean Squared Error (MSE): 26.0226
R-squared: -2.7140

Process finished with exit code 0
```

Fig.21. XGBoost and DNN Performance Evaluation Results

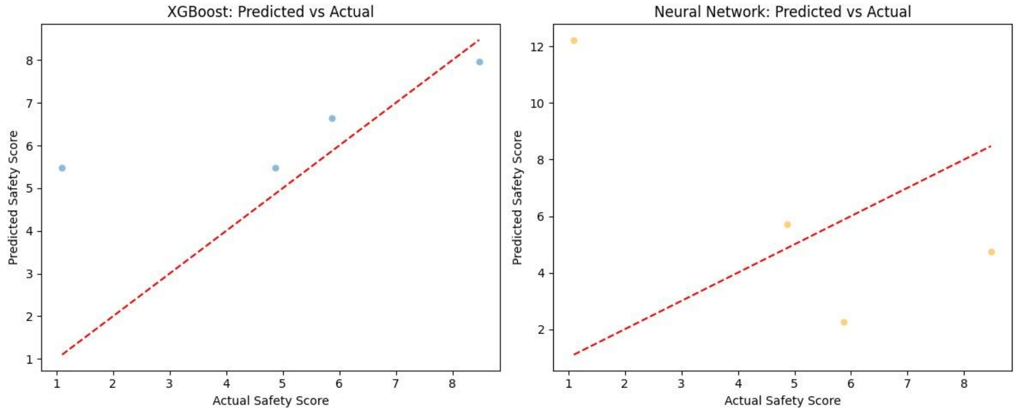
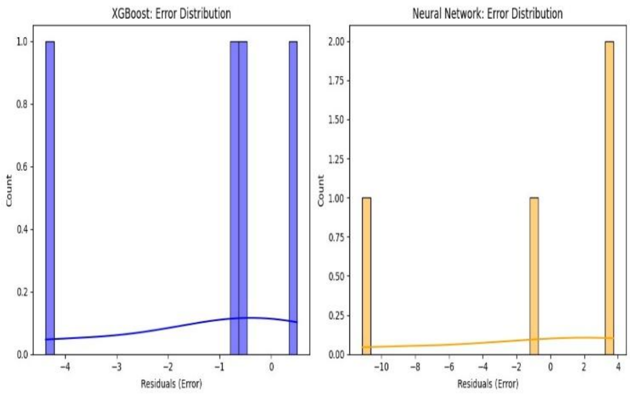


Fig.22. XGBoost and DNN Performance Predicted Vs Actual



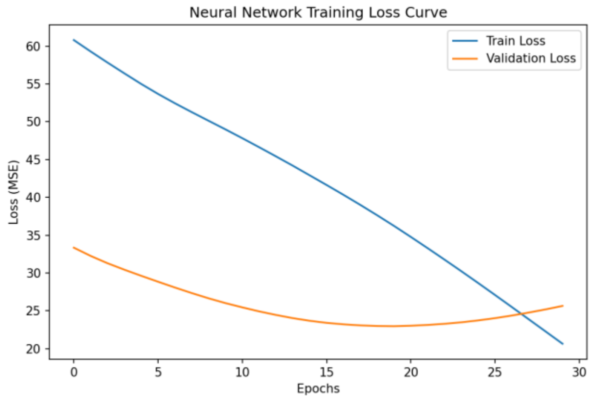


Fig.23. XGBoost and Error Distribution Illustration   Fig.24. XGBoost and DNN Performance Evaluation Result

The XGBoost model had a Mean Squared Error of 5.0973 and an R^2 of 0.2725, which is moderate predictive power with room for improvement but still some level of significant variance explanation. The XGBoost predicted vs. actual safety score plot showed that the predictions followed very closely along the reference diagonal line, which suggests a generally good fit between predicted and observed values. The error distribution plot indicates a relatively narrow spread of residuals, which in turn confirmed the model's robustness across various observations.

Conversely, the model produced by the neural network produced a significantly higher MSE of 26.0226 and a negative R^2 of -2.7140. The negative R^2 is an excellent indication of poor model performance, which can be interpreted to mean that the predictions made by the neural network were poorer than making predictions randomly based on the mean of the target variable. The prediction accuracy plot graphically proves this result, as the large deviations from the line of ideal fit illustrate. The neural network residual plot also shows erratic behavior, with high dispersion and asymmetrical spread of errors, which means that the model failed to generalize on the test set.

In short, on both statistic measures and graphic diagnosis, XGBoost model significantly surpasses the neural network in safety score prediction. Low error rates alongside more consistent predicting patterns indicate XGBoost fits better in this tabular and structured data challenge, particularly for situations where the size of the data and features' complexity is small. The weights used on the safety score formula was chosen based on domain expertise, placing more importance on influential variables like overall crime and violent crime and balancing with favourable variables like tourism and accommodation ratings.Machine learning algorithms were then used to verify if the weighted relationship could be learned from the data and to look for potential nonlinear interactions not defined in the formula. This mixed approach combines the explainability of a structured scoring system with the flexibility and learning capacity of predictive models, ensuring both transparency and accuracy in safety evaluation.

11.Dashboard Implementation

Fig.25. Implemented Safety Tourism DashBoard

To act as the finale of the final phase of the project, Microsoft Power BI was employed to construct an interactive dashboard to efficiently look at and communicate insights gained from the analytical and machine learning portions of the research. The dashboard had been constructed with the goal of enabling straightforward user exploration of safety data on city levels, crime patterns, review sentiments, and model-provided forecasts, to serve as a decision-support tool for visitors, authorities, and policy actors. Various datasets were preprocessed and uploaded to Power BI in CSV form. Slicers and Filters enables visualizations can be filtered by city, type of crime, time of day, year, age group, and gender. This enables in-depth exploration and comparison along various dimensions of the data.

Crime Analytics included:

- Monthly crime distributions and analysis by age and gender are depicted through bar and pie charts.
- Year-wise crime trend lines differentiate between high-, medium-, and low-risk categories of crime based on XGBoost predictions.
- A stand-alone seasonal crime trend visualization specifies differences in crime by different times of the year (e.g., monsoon versus winter).
- Risk-Level Map Visualization: A heat map using a choropleth-style shows clusters of cities using color-coded markers (high risk = red, low = green, medium = orange). This is output of the K-Means clustering on normalized risk features.

- Review-Based Sentiment Analysis: Indicators such as the average rating of tourist attractions and hotels are displayed through gauges. Additionally, bar charts quantify crime and scam mentions in hotel and tourist reviews, as implied through BERT-based sentiment classification.
- Safety Score Prediction: A city ranking bar chart shows cities ranked by their predicted Overall Safety Scores, originally computed using a custom weighted formula of tourism, crime, and accommodation data, and subsequently cross-validated with XGBoost regression. This gives a relative ordering of safety across city destinations.

The dashboard closes the interpretability gap between sophisticated machine learning results and actual-world understanding. It offers a unified, graphical interface to measure regional safety and make travel decisions based on this. By applying spatial (map-based) and temporal (annual trend) analysis, the system is intuitive while data-driven, marrying academic scholarship with real-world usability.

12. Demonstration Of Dashboard Functionalities

2025-05-04 16-54-32.mp4

If you find any difficulties to open the video Click the following link

2025-05-04 16-54-32.mp4

13. Risk Management

During the development and implementation of the project, several risk types were revealed and tackled proactively to ensure project stability and ethics. Loss or corruption of data during merging and preprocessed data handling was prevented through the use of version-controlled copies and exception-handling practices on code scripts. Machine learning model overfitting, especially among the BERT-based classifier model and XGBoost forecaster, was avoided with validation splits, monitoring training/validation loss plot, and cross-validation. For future real-time integration, future API constraints such as rate throttling, down time, and data inconsistency were recognized as threats, especially if derived from third-party crime or tourism data streams. Moreover, synthetic data generation using large language models raised ethical concerns over realism, representation, and potential bias. These were mitigated through the use of structured prompt design, class balanced generation, and post-generation verification of fluency, diversity, and semantic accuracy. Risk awareness in every technical, operational, and ethical aspect was responsible for a more robust and responsible process of development

14. Limitations and Future Enhancement

One significant limitation in this project is data granularity level. City-level analyses and forecasting were done, and that doesn't cover localized variations in safety in a given tourist area or even a neighbourhood. This is due to the unavailability of public crime data at sub-city or location-specific levels. Consequently, the dashboard metrics represent general trends and not hyper-local safety data that would be more applicable to visitors. The other substantial limitation is about the language domain of the sentiment analysis model. The BERT-based classifier used in this project (bert-base-uncased) is only trained on English-language data, and that renders it less effective in multilingual nations like India, where hotel and tourist reviews include Hindi local languages, or code-mixed text. This is limiting the model's inclusiveness and cultural responsiveness. The safety score prediction was based on a weighted formula with normalized crime indicator measures, tourist satisfaction, and accommodation reviews. While the reasoning for the scoring is logical, it lacks empirical evidence from user perception data or official standards. The use of XGBoost and deep learning models was done as a tool to check for internal consistency within the formula and not to produce alternative scoring results. The third interactive dashboard, which is built using Power BI, well depicts city-level aggregated crime and sentiment data. But it relies on static data now and does not incorporate real-time integration or GIS-based mapping of risk areas. Geographic Information Systems (GIS) are essential to contemporary crime analysis because they provide spatial insights that help with reaction and preventative tactics. Designing interventions and identifying hotspots more precisely is made possible by an awareness of neighbourhood-level crime trends[10]. This geographical granularity is also helpful in the tourist context for improving public safety information and assisting visitors in avoiding high-risk regions.

With regards to future enhancement, several changes can strengthen the system and feasible to implement in real life. Firstly, the dashboard can be created to support real-time updates through the inclusion of public safety APIs or open data portals so that the system will be able to offer more dynamic and up-to-date risk intelligence. GIS-based mapping can be added to overlay risks at the neighbourhood or tourist destination level. To increase access and impact, the dashboard can be added to tourism websites, linked to sites like TripAdvisor or OYO, or developed as a mobile application for use by tourists during travel. Multilingual NLP can also be integrated to process regional and code-mixed reviews to make it more inclusive. Lastly, adding features like user-generated alerts, personalized safety recommendations, or travel warnings by category could transform the platform from a passive visualization tool to a clever, context-aware safety guide for tourists and urban planners.

15. Conclusion

The project aimed to bridge the gap in tourism security by developing an integrated, data-driven platform that integrates crime data, user reviews of hotels and tourist destinations, and sentiment analysis. By merging and pre-processing disparate unrelated data sets, the study was able to marry structured crime reports and unstructured review data and offer a multi-faceted perspective of city safety in Indian cities. Sentiment analysis through a fine-tuned BERT model enabled review classification to important risk-related categories, while exploratory data analysis and normality testing supported the formulation of robust feature engineering approaches. Varying machine learning models were implemented for varied prediction tasks. Linear Regression and XGBoost were used to perform city-scale crime trend predictions, while sentiment classification was best with a BERT-based classifier. A module was suggested to predict a safety score to experimentally verify stability for a domain-knowledge-based scoring formula via regression models. Stable performance in the task by XGBoost provided further explanation on the developed safety score reasoning. The final output of a Power BI dashboard is critical information displayed in the form of an interactive dashboard from which users can navigate safety scenarios by city, crime type, and sentiment score. Though there is real-world applicability with the system regarding tourism, local governments, and smart city planners, there are certain limitations with the data granularity and language support. This system has the potential to support safer travel decisions, inform city planners, and serve as a prototype for smart tourism safety platforms

16. Dataset Link

https://www.kaggle.com/datasets/sudhanvahg/indian-crimes-dataset/data

https://www.kaggle.com/datasets/ritvik1909/indian-places-to-visit-reviews-data https://www.kaggle.com/datasets/deeppatel9095/oyo-reviews-dataset

17. Tools and Programming Languages Used

Rstudio

PyCharm

Visual Studio Code

Google Colab

Power BI

18. 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, doi: https://doi.org/10.1016/s0261-5177(03)00003-7.

[2]G. Fuchs and A. Reichel, "An exploratory inquiry into destination risk perceptions and risk reduction strategies of first time vs. repeat visitors to a highly volatile destination," *Tourism Management*, vol. 32, no. 2, pp. 266–276, Apr. 2011, doi: https://doi.org/10.1016/j.tourman.2010.01.012.

[3]M. Nilashi, O. Ibrahim, E. Yadegaridehkordi, S. Samad, E. Akbari, and A. Alizadeh, "Travelers decision making using online review in social network sites: A case on TripAdvisor," Journal of Computational Science, vol. 28, pp. 168–179, Sep. 2018, doi: https://doi.org/10.1016/j.jocs.2018.09.006.

[4]U. Gretzel, M. Sigala, Z. Xiang, and C. Koo, "Smart tourism: foundations and developments," Electronic Markets, vol. 25, no. 3, pp. 179–188, Aug. 2015, doi: https://doi.org/10.1007/s12525-015-0196-8.

[5] N. Levine, "The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report: Document Title: CrimeStat IV: A Spatial Statistics Program for the Analysis of Crime Incident Locations, Version 4.0," 2013. Accessed: Apr. 30, 2025. [Online]. Available: https://www.ojp.gov/sites/g/files/xyckuh241/files/media/document/242960-242995.pdf

[6]Y. Wen, Y. Liang, and X. Zhu, "Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China," PLOS ONE, vol. 18, no. 3, p. e0275382, Mar. 2023, doi: https://doi.org/10.1371/journal.pone.0275382.

[7] I. Bouabdallaoui, F. Guerouate, S. Bouhaddour, C. Saadi, and M. Sbihi, "Sentiment analysis in Tourism: Fine-tuning BERT or sentence embeddings concatenation?," arXiv.org, 2023. https://arxiv.org/abs/2312.07797 (accessed Apr. 28, 2025).

[8] W. Cunha, L. Rocha, and M. A. Gonçalves, "A thorough benchmark of automatic text classification: From traditional approaches to large language models," *arXiv.org*, 2025. https://arxiv.org/abs/2504.01930 (accessed Apr. 29, 2025).

[9] T. Chen and C. Guestrin, "XGBoost: a Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, vol. 1, no. 1, pp. 785–794, Aug. 2016, doi: https://doi.org/10.1145/2939672.2939785.

[10] J. Kurland and S. D. Johnson, "The Influence of Stadia and the Built Environment on the Spatial Distribution of Crime," *Journal of Quantitative Criminology*, vol. 37, no. 2, pp. 573–604, Dec. 2019, doi: https://doi.org/10.1007/s10940-019-09440-x.