PCMJ

Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

VISUALIZING TOURIST BEHAVIOUR IN MELAKA USING HDBSCAN AND MULTIPLE LINEAR REGRESSION

MOHAMAD HALID SAZLI HAMSAH
Faculty of Computing and Mathematical Sciences, Universiti Teknologi MARA
2023305247@student.uitm.edu.my

MOHD TAUFIK BIN MISHAN*
Faculty of Computing and Mathematical Sciences, Universiti Teknologi MARA
mtaufik@uitm.edu.my

Article Info

Abstract

This research offers a visual solution for patternizing Melaka's tourist attractions based on Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and Multiple Linear Regression (MLR). Conventional techniques for patternizing tourists' behavior, e.g., surveys, are time-consuming with a limited sample size. To address this, the current research utilizes geo-tagged photo data and past tourist arrival data to patternize travel behavior and forecast tourist arrival. HDBSCAN clusters tourist attractions effectively, revealing areas of density, and MLR accurately predicts tourist visits according to influence factors. The system is realized as a web application, offering interactive heatmaps and trend analysis using visualization methods like line charts, bar charts, pie charts, and scatter plots. The results of the usability test show high satisfaction, which reflects the system's potential for supporting tourism agencies and businesses in making decisions. This combination of machine learning and data visualization enables more effective, data-driven management of tourism in Melaka.

Received: March 2025
Accepted: September 2025
Available Online: November 2025

Keywords: Tourist Local Attraction, Tourist Behavior, Machine Learning, HDBSCAN, Multiple Linear Regression, Data Visualization

INTRODUCTION

Tourism is a key driver of economic growth, job creation, and cultural exchange globally. As an essential sector, tourism contributes significantly to the economy, with many countries relying on its revenue for sustainable development (Keertika et al., 2021). In Malaysia, Melaka is one of the most visited historical cities, attracting both domestic and international tourists. However, rapid tourism growth has led to challenges such as overcrowding, resource misallocation, and inconsistent visitor experiences. Understanding tourist behavior is crucial for optimizing tourism management and ensuring sustainable tourism development (Arthan et al., 2021).

Traditional data collection methods, such as surveys and manual observations, have limitations in terms of sample size, time constraints, and subjectivity. The rise of digital technologies, particularly social media and geospatial data, provides new opportunities to analyze tourist behavior in a more efficient and data-driven manner (Jaitrong & Nanthaamornphong, 2024). Machine learning techniques, such as clustering and regression, have been widely used in various industries to identify patterns and predict trends. In this study, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is utilized to cluster tourist activity hotspots, while Multiple Linear Regression (MLR) is employed to predict future tourist arrivals.

This research aims to address overtourism concerns, improve tourism management, and enhance visitor experiences through data-driven insights. By leveraging geo-tagged photo data and historical tourist arrival statistics, this study develops a web-based system that visualizes tourist behavior and predict future trends. The system incorporates interactive visualizations such as heatmaps, line charts, and dashboards to present actionable insights for tourism agencies and businesses. Ultimately, integrating machine learning with data visualization enhances decision-making, providing stakeholders with a comprehensive understanding of tourist movements and preferences (Rialti et al., 2019).

.

LITERATURE REVIEW

Tourist behavior analysis has been widely studied in the field of tourism and data science. Understanding the factors influencing tourist movement patterns and preferences is crucial for sustainable tourism development (Beritelli et al., 2021). Various factors, including cultural background, personal motivations, and external influences such as economic and political conditions, play a role in shaping tourist decisions. Traditionally, researchers have relied on surveys and observational studies to gather insights into tourist behavior, but these methods are often constrained by sample size limitations and time-intensive data collection processes (Ghosh, 2023).

Machine Learning for Tourist Analysis

Machine learning has revolutionized the way researchers analyze large-scale tourism datasets. Clustering algorithms, such as K-means, DBSCAN, and HDBSCAN, have been widely used to classify tourist movement patterns and identify hotspots. K-means clustering is effective for simple datasets but struggles with non-spherical clusters and varying densities (Kingrat, 2023). DBSCAN improves upon K-means by detecting clusters of arbitrary shape but requires careful parameter tuning. HDBSCAN further refines this by automatically detecting the optimal number of clusters and filtering out noise, making it particularly suitable for geospatial tourist data (McInnes et al., 2017). Studies have demonstrated that HDBSCAN outperforms other clustering algorithms in detecting complex tourist movement patterns and seasonal trends (Sahu et al., 2024).

The low quality of some clusters, in which some clustered points appear to be quite far from the main aggregation and should be regarded as noise, is especially concerning. Fortunately, an outlier score is used by the HDBSCAN algorithm to provide a solution. This score goes from zero to one, where a value close to one indicates the opposite and zero indicates a high degree of confidence that the point is part of the cluster. Figure 1 shows the result of HDBSCAN algorithm.

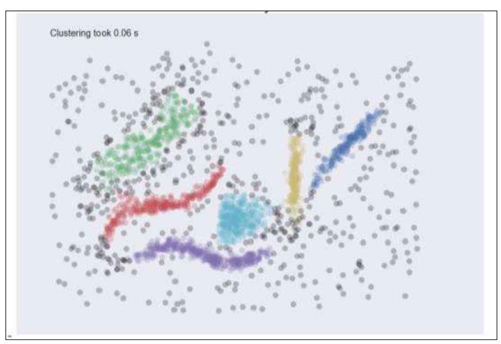

Figure 1: Clusters found by HDBSCAN

Predictive Modeling in Tourism

Predictive analytics plays an essential role in forecasting tourist arrivals and planning resource allocation. Multiple Linear Regression (MLR) has been extensively used to predict tourism demand based on historical data, economic indicators, and seasonality trends (Tranmer et al., 2020). The accuracy of MLR models depends on selecting relevant independent variables, such as exchange rates, GDP growth, and social media trends, which influence tourist arrivals. Research has shown that well-calibrated MLR models can provide reliable forecasts, helping tourism agencies optimize marketing strategies and infrastructure planning (Amar & Mgt, 2021). Combining MLR with machine learning techniques such as decision trees and neural networks further enhances predictive accuracy, allowing for dynamic forecasting in rapidly changing tourism environments (Manimala et al., 2023). The multiple linear regression equation follows the same structure as the simple linear regression equation but includes additional terms as shown in equation (1).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x2_i + \cdots + \beta_p x_{pi} + e_i \qquad \mathbf{1}$$

Visualization Techniques in Tourism Research

Effective visualization techniques are crucial for transforming raw data into actionable insights. Heatmaps are widely used to depict high-density tourist activity areas, allowing policymakers to identify overcrowding issues and improve visitor management (Guo et al., 2020). Line charts and bar graphs provide historical trend analysis, helping businesses and tourism agencies understand seasonality patterns. Interactive dashboards with real-time analytics enable stakeholders to monitor visitor movement and optimize service offerings. Studies highlight that incorporating intuitive visualizations into decision-making processes enhances the efficiency of tourism management, ensuring a better experience for both tourists and local communities (Yang et al., 2021). Recent advancements in geospatial visualization techniques, such as 3D mapping and augmented reality overlays, offer new possibilities for engaging users with immersive tourism insights. As shown in Figure 2, DV encompasses several processes, such as data collection, cleaning and preparation, analysis, graphical interpretation, and implementation and deployment (Parul Gandhi, 2021).

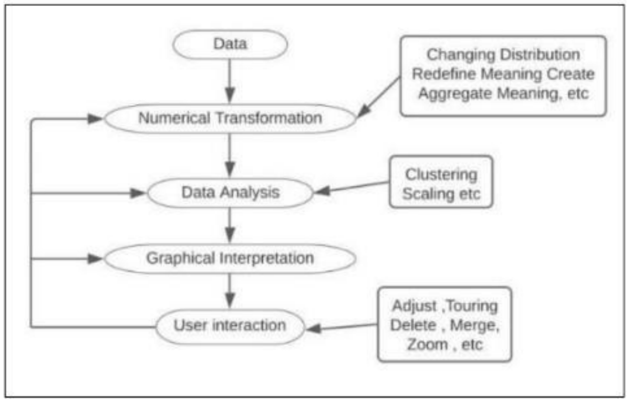

Figure 2: Flow of Data Visualization

METHODOLOGY

Understanding tourist local attraction requires a structured approach that effectively analyzes movement patterns, clusters significant locations, and predicts future trends. This study employs a combination of clustering and regression techniques to gain insights into how tourists interact with different locations in Melaka. By leveraging large datasets from social media and official tourism records, this research ensures a data-driven methodology that enhances the accuracy and reliability of the findings. The methodology follows a structured process that includes data collection, preprocessing, clustering analysis, predictive modeling, system development, and testing.

Research Design

This study adopts a modified waterfall methodology, ensuring a structured and sequential approach to system development. The methodology consists of several key phases: data collection, preprocessing, clustering analysis, predictive modeling, system development, and testing. Each phase was carefully designed to maximize efficiency and ensure that the results accurately reflect tourist behavior patterns in Melaka. The waterfall methodology ensures that each stage is completed before proceeding to the next, minimizing errors and ensuring a robust final implementation.

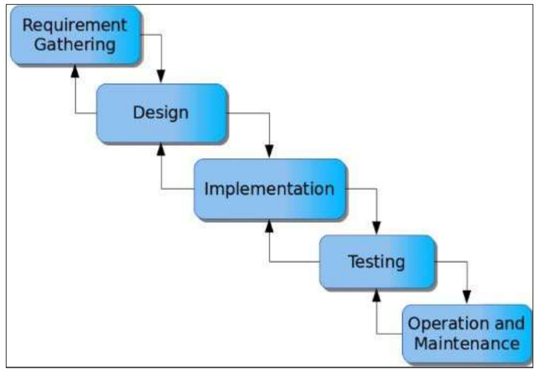

Figure 3: Modified Waterfall Methodology

Data Collection

Data were collected from multiple sources to ensure comprehensive coverage of tourist movement and behavior. Geo-tagged photos were extracted from social media platforms such as Instagram and Flickr, providing insights into popular tourist attractions and movement patterns. Additionally, official statistics on tourist arrivals and demographics were obtained from government databases to complement the social media data. To account for external factors influencing tourist behavior, weather data and economic indicators were collected from meteorological services and financial reports. This multi-source approach ensures a diverse and comprehensive dataset, capturing different aspects of tourism trends and movement patterns.

The data utilized in this project is primary data gathered through the Flickr API to retrieve geotagged photographs within Melaka's geographical boundary. Data scraping was done through a Python script that extracted photo metadata, such as latitude and longitude coordinates, for cluster analysis. The bounding box technique was applied to specify the area of interest so that only Melaka data was scraped. 35,021 successfully geotagged data points were collected for analysis. Figure 4 shows the example of the dataset.

Progress in Computer and Mathematics Journal (PCMJ)
volume 3 [November, 2025]
e-ISSN: 3030-6728
Website: fskmjebat.uitm.edu.my/pcmj

	A	B	C
1	id	latitude	longitude
2	2.33E+38	2.203039	102.2503
3	1.12E+37	2.1927	102.2492
4	6.2E+37	2.194447	102.2487
5	2.14E+38	2.195383	102.2505
6	3.32E+38	2.194572	102.2486
7	1.28E+38	2.194777	102.2486
8	2.78E+38	2.192672	102.2496
9	3.15E+38	2.192647	102.2496
10	6.59E+37	2.191719	102.2502
11	3.24E+38	2.192755	102.2492
12	1.23E+38	2.196943	102.2464
13	2.77E+38	2.196943	102.2464
14	2.67E+38	2.194864	102.2492
15	3.04E+38	2.196943	102.2464

Figure 4: Geotagged Dataset in Melaka from Flickr

A secondary data set as shown in Figure 5 was gathered from the Malaysia government website containing official statistics of tourist arrivals in Melaka. The data set is for the year 2000 to 2019 and contains different attributes like year, Average Length of Stay, Tourist Arrivals (Millions), Tourist Receipts (RM'000) and RM Spend/Day. The data set was utilized to create a multiple linear regression model which was capable of predicting tourist arrivals based on different factors. The data contain pertinent attributes that influence tourist inflow and are a critical part of predicting future tourist arrivals.

	A	B	C	D	E
1	YEAR	AVG LENGTH OF STAY	TOURIST ARRIVALS (MILLIONS)	TOURIST RECEIPTS (RM'000)	RM SPEND/ DAY
2	2000	1.81	2.17	1,148.46	292.4
3	2001	1.98	2.57	1,581.54	310.8
4	2002	2.02	2.98	1,497.07	248.703
5	2003	1.34	3.6	1,348.80	279.499
6	2004	1.58	4.01	1,994.32	314.771
7	2005	1.78	4.69	2,661.66	318.83
8	2006	1.86	5.1	3,161.32	333.403
9	2007	1.87	6	3,172.05	281.62
10	2008	1.88	7.21	3,853.66	284.48
11	2009	1.89	8.91	4,481.75	266.28
12	2010	2.13	10.35	6,331.89	287.09
13	2011	2.15	12.17	7,641.05	294.87
14	2012	2.25	13.71	9,107.55	295.22
15	2013	2.28	14.31	10,997.32	337
16	2014	2.01	15.03	12,013.24	397.6
17	2015	2.18	15.74	16,759,75	488.53
18	2016	2.16	16.28	18,289.49	520.04
19	2017	2.25	16.79	19,651.04	445.9
20	2018	2.46	17.02	20,979.52	501.07
21	2019	2.55	18.73	21,298.60	446

Figure 5: Melaka Tourist Arrivals Dataset

Data Preprocessing

Data preprocessing involved cleaning and readying datasets for multiple linear regression models and clustering. In the case of the clustering model as shown in figure 6, it loaded the Flickr Geotagged Data (Melaka) dataset through Pandas and initially explored it to discover its structure, identify missing values, and discover inconsistencies. It deleted missing values in latitude and longitude columns for proper geospatial clustering and removed duplicate records to prevent redundancy. Latitude and longitude values were specifically cast to float data types for uniformity, and leading and trailing whitespaces from string columns were stripped to avoid formatting problems. Validation checks kept latitude and longitude values within valid ranges, and geotagged data out of range was removed. Having performed these preprocessing tasks, the dataset was exported as a cleaned CSV file for subsequent clustering analysis.

```python
# Step 3: Handle Missing Values
data.dropna(subset=['latitude', 'longitude'], inplace=True)

# Step 4: Remove Duplicates
data.drop_duplicates(inplace=True)

# Step 5: Standardize Data Types
data['latitude'] = data['latitude'].astype(float)
data['longitude'] = data['longitude'].astype(float)

# Step 6: Trim Whitespace
# Apply trimming to all string columns
string_columns = data.select_dtypes(include=['object']).columns
data[string_columns] = data[string_columns].apply(lambda x: x.str.strip())

# Step 7: Validate Geotagged Data
# Check if latitude and longitude are within valid ranges
data = data[(data['latitude'] >= -90) & (data['latitude'] <= 90)]
data = data[(data['longitude'] >= -180) & (data['longitude'] <= 180)]
```

Figure 6: Data Cleaning and Validation for Clustering Model

Likewise, preprocessing was carried out for the multiple linear regression model with the Melaka Tourist Arrivals dataset to clean and normalize the data. The dataset was imported with correct encoding and numerical integrity preserved by converting tourist receipts values, which were initially formatted with comma separation, to a float data type for numerical computation as shown in figure 7. Missing values were inspected, and column names were standardized by substituting spaces and special characters with underscores and to lowercase for better accessibility and consistency in data handling as shown in Figure 8. The cleaned dataset was saved as a CSV file for regression modelling as shown in Figure 9. These preprocessing measures ensured that both datasets were neatly structured, free from irregularities, and prepared for use in their respective models.

```
# Remove commas and convert to float
data["TOURIST RECEIPTS (RM'000)"] = data["TOURIST RECEIPTS (RM'000)"].replace(',', '', regex=True).astype(float)
```

Figure 7: converting tourist receipt values into a float data type for numerical analysis

```
# Standardize column names
data.columns = data.columns.str.replace('[^\w\s]', '', regex=True).str.replace(' ', '_').str.lower()
```

Figure 8: standardized column names by replacing spaces and special characters with underscores

```
   year  avg_length_of_stay  tourist_arrivals_millions  \
0  2000                1.81                       2.17
1  2001                1.98                       2.57
2  2002                2.02                       2.98
3  2003                1.34                       3.60
4  2004                1.58                       4.01

   tourist_receipts_rm000  rm_spend_day
0                 1148.46       292.400
1                 1581.54       310.800
2                 1497.07       248.703
3                 1348.80       279.499
4                 1994.32       314.771
missing value:
YEAR                          0
AVG LENGTH OF STAY            0
TOURIST ARRIVALS (MILLIONS)   0
TOURIST RECEIPTS (RM'000)     0
RM SPEND/ DAY                 0
```

Figure 9: Dataset for regression after data-preprocessing

Clustering Analysis Using HDBSCAN

HDBSCAN was employed to identify high-density tourist activity zones. The clustering process involved optimizing parameters such as minimum cluster size and distance thresholds to refine clustering accuracy. Outliers and noise were filtered out to ensure the focus remained on meaningful tourist hotspots. The resulting clusters were validated by comparing them with historical data and known tourist attraction locations, ensuring that the model accurately captured real-world movement patterns. HDBSCAN's hierarchical approach allows it to adapt to varying densities of tourist visits, making it more effective than traditional clustering methods such as K-means, particularly in complex environments like Melaka's diverse tourist sites.

Predictive Modeling Using Multiple Linear Regression (MLR)

MLR was implemented to forecast tourist arrivals based on historical trends and external influencing factors. The predictive modeling process began with feature selection, identifying key variables such as past visit trends, economic indicators, and weather conditions that significantly impact tourist numbers. The model was then trained using historical data, optimizing regression coefficients to enhance predictive accuracy. Performance evaluation was conducted using validation metrics such as R-squared and Mean Absolute Error (MAE), ensuring the reliability of the forecasts. Furthermore, cross-validation techniques were applied to prevent overfitting, ensuring that the model remains robust in predicting future trends under varying conditions.

System Development

A web-based application was developed to visualize the analyzed data and provide stakeholders with interactive insights. The system features heatmaps that display high-density tourist areas based on clustering analysis, allowing users to identify popular attractions. Trend charts were integrated to show tourist arrival patterns over time, helping stakeholders plan for seasonal fluctuations. Additionally, interactive dashboards were developed to enable users to explore predictive models, adjust parameters dynamically, and extract meaningful insights to support decision-making. The user interface was designed with a focus on simplicity and usability, ensuring that stakeholders from various backgrounds can access and interpret the data with ease.

Testing and Validation

The system underwent rigorous testing to ensure functionality, accuracy, and usability. Accuracy testing was conducted by comparing predicted tourist arrivals with actual records, validating the effectiveness of the MLR model. Functionality testing evaluated the responsiveness and performance of the web application, ensuring seamless navigation and data representation. Finally, usability assessment was carried out by gathering feedback from tourism stakeholders, leading to refinements in the user interface and system functionalities to enhance user experience and decision-making tools. Stress testing was also conducted to evaluate the system's performance under high user loads, ensuring that it remains stable during peak usage times.

RESULT AND DISCUSSION

This section presents the findings of the study, focusing on the analysis of tourist behavior in Melaka. The results provide insights into the clustering of tourist activity, the accuracy of the predictive model, and the overall performance of the developed visualization system. By integrating machine learning techniques, this research aims to enhance tourism management through data-driven decision-making. The following subsections discuss the key findings in detail, including the clustering analysis, predictive model performance, system usability, and practical implications for tourism stakeholders.

Clustering Analysis Using HDBSCAN

Figure 10 illustrates the HDBSCAN clustering result of tourist activity in Melaka, where each purple dot represents a geotagged tourist location in terms of latitude and longitude. The clustering algorithm highlights areas of high density in cyan and blue, indicating popular tourist spots. The dense tourist activity is observed along latitude ~2.20 to 2.25 and longitude ~102.20 to 102.30, which corresponds to well-known attractions such as Jonker Street, A'Famosa, and Stadthuys. In contrast, the scattered purple dots on the map represent outliers or less frequently visited locations, indicating regions with lower tourist density.

As shown in Figure 10 there were 484 clusters formed during the HDBSCAN clustering process, representing unique tourist activity regions. The clusters reflect several patterns of tourist movement, with some regions more dense than others. Also, 629 data points (1.80% of the data) were identified as noise and not included in the clustering. The noise points indicate regions of low data density, indicating regions of small or no tourist activity, or outliers in the data. Noise points can be explained by occasional or rare visits and are therefore less important to be part of a significant cluster.

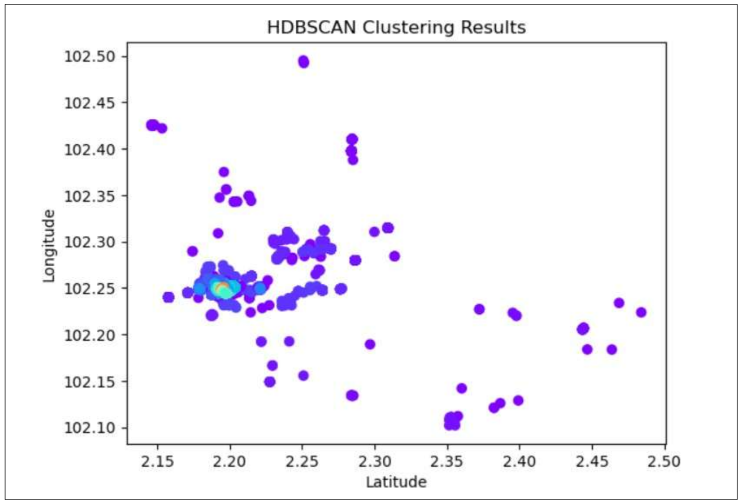

Figure 10: Visualization of HDBSCAN Clustering Results

Figure 11 illustrate the heatmap portrays tourist flow patterns, where red and yellow areas are high-density areas and green and blue areas are low-density areas. The highest concentration of tourist activity is within Ayer Keroh and Melaka city center, covering key attractions like Melaka City, Jonker Street, A'Famosa, and the Melaka River Cruise. On the other hand, regions like Jasin, Alor Gajah, and Masjid Tanah experience moderate tourist activity, which is to say that they are either secondary destinations or less visited areas.

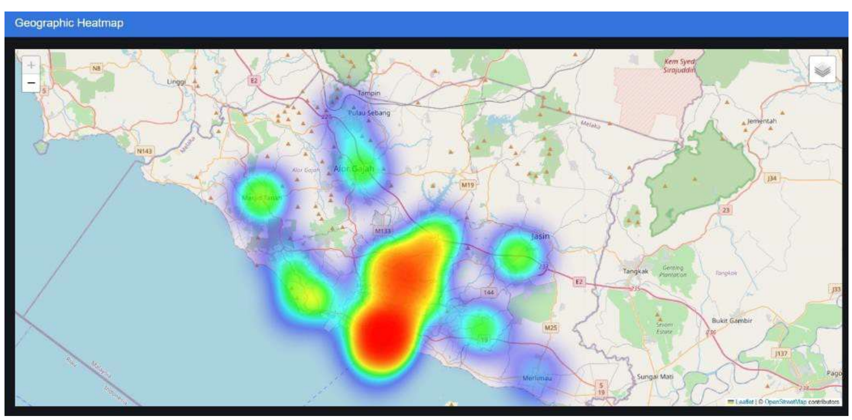

Figure 11: Heatmap visualization based on the Clustered Data

Predictive Accuracy of Multiple Linear Regression (MLR)

The multiple linear regression (MLR) model was developed to predict tourist arrivals from the significant determinant variables, which were the year, length of stay average, and RM spend per day. The model was a very good predictor with an average R-squared of 0.96, which meant that 96% of the variations in tourist arrivals were explained by the variables. With a 5-Fold Cross-Validation technique and 80:20 split, the model had an average Mean Absolute Error (MAE) of 1.11 and an average Root Mean Squared Error (RMSE) of 1.04, indicating low prediction error as shown in Figure . With training on the entire dataset, the model performed even better with an R-squared of 0.98, MAE of 0.73, and RMSE of 0.84. These findings confirm the model's effectiveness in estimating tourism trends in statistical data, which highlights the effectiveness of variables like the tourist spending pattern and length of stay in predicting visitor arrivals. These findings contribute to essential input for Melaka tourism decision and planning.

```
Cross-Validation RMSE Scores: [1.23213264 0.88848348 0.85348269 0.92463592 1.29641398]
Cross-Validation MAE Scores: [1.51815084 0.78940289 0.7284327  0.85495159 1.68068922]
Cross-Validation R² Scores: [0.96641948 0.95956294 0.96916735 0.94155476 0.95307582]
Average RMSE: 1.039
Average MAE: 1.114
Average R²: 0.958
```

Figure 12: 5-Fold Cross-Validation approach (80:20 split) Result

The plot of figure 13 shows the actual and predicted tourist arrivals across time from the multiple linear regression (MLR) model. Observed actual tourist arrivals are shown in blue, whereas the predicted values based on the model are shown in red. The tight follow-up of the two lines confirms the model's high accuracy in capturing trends, which is reinforced by an R-squared measure of 0.96 when cross-validated and 0.98 when trained with all the data.

In addition, the dashed red line extrapolates future tourist arrivals outside the available data, demonstrating an upward trend. The projection reveals the value of the model in creating vital information for future planning in the tourism sector. The slight differences between predicted and actual values validate that the year, length of stay, and RM spending per day are influential factors in tourist arrivals. Generally, the plot vouches for the validity and possible application of the model in decision-making in the tourism sector of Melaka.

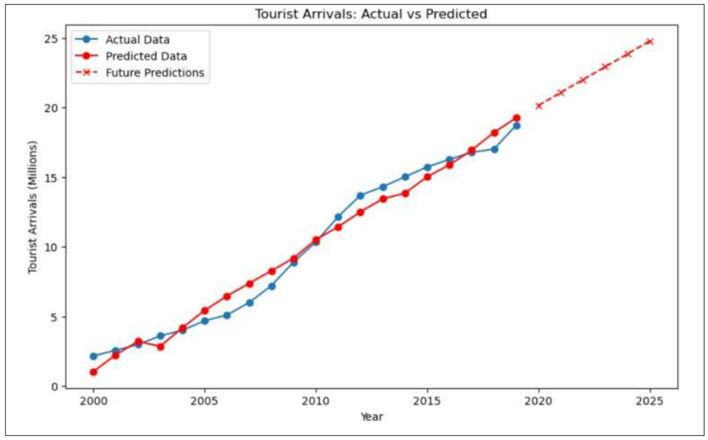

Figure 13: Comparison Between Actual and Predicted Tourist Arrivals Over Time

Visualization and System Performance

The developed web-based system effectively presented analytical results through interactive visualizations, including heatmaps, trend charts, and dashboards. Users found the heatmaps particularly useful for identifying high-density tourist areas in real time, while the trend charts provided valuable insights into long-term visitation patterns. The system's responsiveness and user-friendly interface were positively evaluated during functionality testing. The real-time data integration feature was well-received, as it allowed stakeholders to make data-driven decisions quickly. Overall, the system's performance was deemed efficient, and its usability was validated through user feedback.

Implications and Recommendations

The findings of this study have several practical implications for tourism management in Melaka. The identification of tourist hotspots can aid in better traffic management and resource allocation. The predictive model can assist businesses in forecasting demand, allowing them to optimize staffing and inventory levels accordingly. Furthermore, local authorities can use these insights to implement targeted promotional campaigns during low seasons and adopt sustainable tourism practices to prevent overcrowding in peak seasons. Future improvements to the system could include integrating additional data sources such as live GPS tracking and sentiment analysis from social media reviews to enhance prediction accuracy and visitor experience assessment.

CONCLUSION

This study successfully demonstrated the potential of machine learning techniques in analyzing and predicting tourist local attraction in Melaka. By implementing HDBSCAN for clustering and MLR for predictive modeling, the research provided valuable insights into tourist movement patterns and factors influencing visitor influx. The developed web-based visualization system effectively transformed raw data into actionable intelligence, enabling stakeholders to make informed decisions regarding tourism management and infrastructure planning.

The results highlighted the effectiveness of HDBSCAN in identifying tourist hotspots and the strong predictive capabilities of MLR in forecasting tourist arrivals. The system's usability and performance were validated through testing and feedback, confirming its practicality for real-world applications. Additionally, the study underscored the importance of data-driven decision-making in enhancing visitor experiences and ensuring sustainable tourism development.

Future research could further refine the predictive model by integrating additional factors such as real-time weather updates, social media sentiment analysis, and transportation data. Expanding the dataset to include more diverse tourist demographics could also enhance the accuracy and applicability of the findings. Ultimately, this study contributes to the growing field of intelligent tourism management and serves as a foundation for further advancements in data-driven tourism analytics.

REFERENCES

Adediran, S. S. A. A. (2016). An overview of big data visualization techniques in data mining.

Agrawal, R. (2024). Polynomial regression for beginners. Analytics Vidhya.

Ahmadi, R., Ekbatanifard, G., & Bayat, P. (2021). A modified grey wolf optimizer based data clustering algorithm. Applied Artificial Intelligence, 35(1), 63–79.

Akter, S., Bandara, R., Hani, U., Fosso Wamba, S., Foropon, C., & Papadopoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. International Journal of Information Management, 48, 85–95.

Amar, J. C., & Mgt, D. (2021). Tourists' arrival prediction using regression techniques. International Journal of Science and Research, 10(11), 773–784.

Arthan, S., Jandum, K., & Tamee, K. (2021). Exploring tourist behavior from social media using geotagged photographs. 2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering (ECTI DAMT and NCON 2021), 285–288.

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An introduction to machine learning. Clinical Pharmacology and Therapeutics, 107(4), 871–885.

Beritelli, P., Reinhold, S., & Laesser, C. (2021). Visitor flows, trajectories and corridors: Planning and designing places from the traveler's point of view. Annals of Tourism Research, 82, 102936.

Correia, A., & Dolnicar, S. (2021). Women's voices in tourism research. The University of Queensland.

Ghosh, P. (2023). Understanding tourist behaviour towards destination selection based on social media information: An evaluation using unsupervised clustering algorithms..