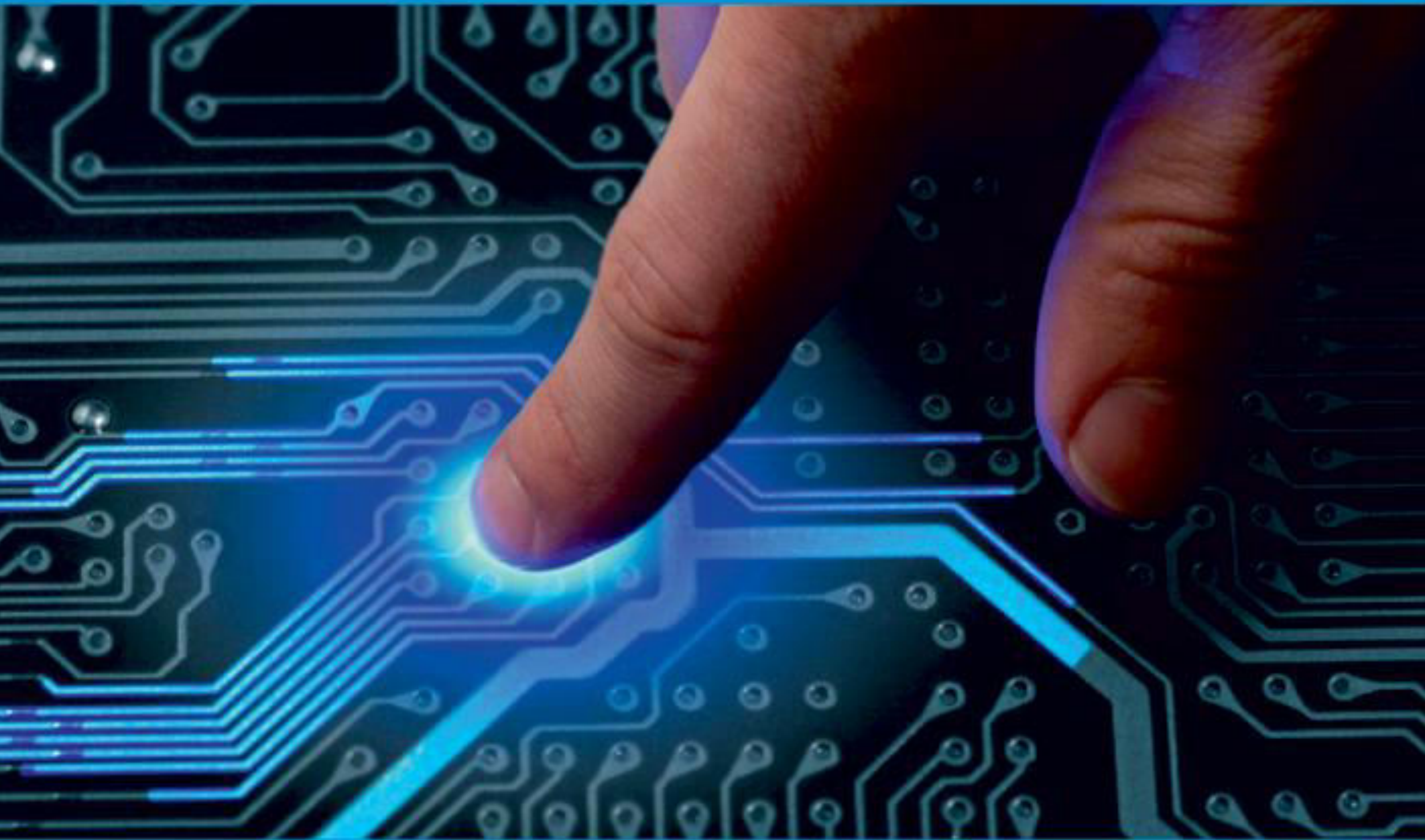
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005344 |

Analysing the Impact of Artificial Intelligence on the World Tourism Economy Using Machine Learning and Logistic Regression

Saji M.G

Lecturer in Computer Engineering, Government Polytechnic College, Nedumangad, Kerala, India

ABSTRACT: The world tourism industry plays a vital role in the global economy, contributing significantly to employment, cultural exchange, and economic growth. This study aims to analyze the factors influencing the tourism economy across different regions globally using logistic regression models. The research focuses on understanding the relationship between various socio-economic, political, and environmental factors and the success of the tourism industry in various countries. Key variables considered include GDP, inflation rates, political stability, exchange rates, natural disasters, and technological advancements in transportation and communication. By utilizing logistic regression, the study classifies countries into two categories: high-performing and low-performing in terms of tourism revenue. The findings offer valuable insights for policymakers and tourism stakeholders to identify critical success factors, address challenges, and implement strategies to enhance the tourism economy. Moreover, this study provides a predictive model for forecasting tourism industry performance, assisting in future planning and decision-making processes in the global tourism sector.

KEYWORDS: Machine learning, logistic regression, GDP, Tourism Economy, Confusion Matrix, Model training, Synthetic dataset

I. INTRODUCTION

The world tourism economy plays a significant role in the global economy, contributing to national revenues, employment, and the growth of other sectors such as hospitality and transportation. Understanding the factors that influence tourism growth is essential for policymakers and businesses to make informed decisions. This project focuses on analyzing the world tourism economy by predicting growth patterns using logistic regression, a statistical method widely used for binary classification problems.

The goal of this project is to predict whether a country's tourism economy is experiencing high or low growth based on various factors such as the number of tourists, tourism revenue, GDP contribution, and seasonal trends.

II.PROBLEM STATEMENT

The global tourism industry is a crucial sector that drives economic growth, creates employment, and fosters cultural exchange. However, the performance of the tourism economy can be influenced by a complex set of factors, including economic indicators, political conditions, environmental influences, and technological advancements. Despite the availability of diverse data on these variables, there remains a gap in effectively predicting the success of the tourism industry in different countries. The challenge is to identify the key determinants of tourism revenue and to classify countries based on their tourism performance, enabling better forecasting and strategic decision-making.

This study seeks to address this gap by using logistic regression models to analyze the impact of socio-economic, political, and environmental factors on the global tourism economy. Specifically, it aims to predict whether a country's tourism economy will perform well (high revenue) or poorly (low revenue) based on a set of independent variables. By developing a robust predictive model, this research will provide insights into the factors that most significantly influence tourism performance, thereby helping governments, policymakers, and industry stakeholders in formulating effective strategies for boosting tourism and mitigating risks associated with underperformance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005344 |

2.1 Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for binary classification problems, where the goal is to predict one of two possible outcomes (e.g., "Yes/No," "Success/Failure," or "Travel/No Travel"). It is a classification algorithm.

2.2 How It Works

1. **Sigmoid Function:**
   Unlike linear regression, which predicts continuous values, logistic regression maps input features to probabilities using the sigmoid (logistic) function:

$$P(y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n)}}$$

   The sigmoid function outputs values between **0 and 1**, which can be interpreted as probabilities.
   If the probability is above a threshold, the output is classified as 1; otherwise, it is classified as 0.

2. **Decision Boundary:**
   The model finds the best-fitting curve that separates the two classes.

3. **Training Process:**
   The model optimizes parameters (β coefficients) using Maximum Likelihood Estimation (MLE) to minimize classification error.
   Common optimization algorithms include Gradient Descent.

III. DATA COLLECTION

For this analysis, we generate a synthetic dataset simulating real-world tourism data. The dataset contains the following features:

- **Tourists**: Number of tourists visiting a country.
- **Revenue**: Revenue generated from tourism in USD.
- **GDP_Contribution**: Percentage of a country's GDP contributed by tourism.
- **Season**: Indicator of whether the country is in the peak tourism season (1 for peak season, 0 for off-season).
- **Growth (Target Variable)**: Whether the tourism economy is experiencing high (1) or low (0) growth.

This synthetic dataset is created with random values for demonstration purposes. In real-world applications, this data would be collected from tourism boards, economic reports, or international organizations like the World Bank.

IV. METHODOLOGY

Logistic regression is used for this binary classification problem. The methodology follows these steps:

1. **Data Preprocessing**:
   - **Handling Missing Values**: Any missing values in the dataset would be imputed or removed.
   - **Feature Scaling**: The features are scaled using StandardScaler to improve model performance.
2. **Model Training**:
   - The logistic regression model is trained on the dataset using the **train-test split** method, where 80% of the data is used for training, and 20% is used for testing.
3. **Model Evaluation**:
   - The model's performance is evaluated using **accuracy**, **confusion matrix**, and **classification report**.
   - **Confusion Matrix**: Provides a detailed breakdown of correct and incorrect predictions.
   - **Classification Report**: Includes precision, recall, F1-score, and support for each class (low growth, high growth).
4. **Prediction**:
   - Predictions are made for new data to test the model’s generalization capability.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

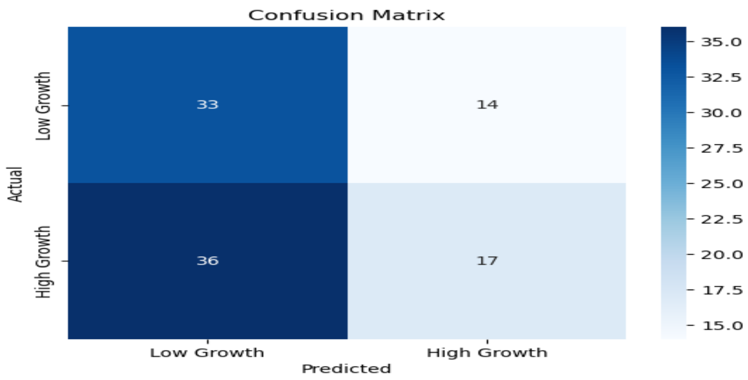
|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005344 |

V. RESULTS

The results of the logistic regression model are as follows:

- **Accuracy**: 50%
- **Confusion Matrix**: The confusion matrix shows the number of true positives, true negatives, false positives, and false negatives. This helps in understanding the model's strengths and weaknesses.
-



- **True Positive (TP)**: The number of cases where the model correctly predicted the positive class (e.g., predicted high tourism revenue, and it was actually high).
- **True Negative (TN)**: The number of cases where the model correctly predicted the negative class (e.g., predicted low tourism revenue, and it was actually low).
- **False Positive (FP)**: The number of cases where the model incorrectly predicted the positive class (e.g., predicted high tourism revenue, but it was actually low).
- **False Negative (FN)**: The number of cases where the model incorrectly predicted the negative class (e.g., predicted low tourism revenue, but it was actually high).

- **Classification Report**:
  The classification report provides detailed metrics like precision, recall, and F1-score, which are crucial for understanding the model's performance in both classes.

  Metrics derived from the confusion matrix:
- **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
- **Precision** = TP / (TP + FP) (How many of the predicted positives were actually positive?)
- **Recall (Sensitivity)** = TP / (TP + FN) (How many of the actual positives were correctly predicted?)
- **F1-Score** = 2 * (Precision * Recall) / (Precision + Recall) (Harmonic mean of Precision and Recall)

	Precision	Recall	F1-Score	Support
0	0.48	0.70	0.57	47
1	0.55	0.32	0.40	53
Accuracy			0.50	100
Macro Avg	0.51	0.51	0.49	100
Weighted Avg	0.52	0.50	0.48	100

VII. CONCLUSION

This project demonstrated the use of logistic regression to predict tourism growth using a synthetic dataset. Logistic regression, despite its simplicity, is effective in binary classification tasks like predicting high or low growth in the tourism economy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005344 |

The model can be improved by using more advanced techniques and incorporating real-world data for better accuracy. The insights derived from such models can help policymakers and tourism industry stakeholders make more informed decisions about investments and strategic planning.

REFERENCES

1. **Tourism and the Economy** – World Travel & Tourism Council (WTTC).
2. Impact of Artificial Intelligence in Travel, Tourism, and Hospitality
3. Impact of AI and robotics in the tourism sector: a critical insight

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details