Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

Research on Juneau Tourism Industry Based on Sustainable Development

Peishuo Si*, Kaiwei Jing, Shuning Lin, Xu Wang

Electronic and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China
* **Corresponding author**: Peishuo Si (Email: 17752134209@163.com)

Abstract: This study focuses on the sustainable tourism development strategies for Juneau's tourism industry. By comprehensively analyzing factors such as the resident population, cruise passenger numbers, and ecological investment, the study aims to optimize tourism development and achieve sustainability in economic, environmental, and social aspects. The research begins with comprehensive data collection for different issues. Given that the collected data had some missing parts, Lagrange interpolation was used to fill in the missing values, and box plots were utilized to detect outliers. Subsequently, scatter plots and pairwise relationship plots were used to explore the data to observe its intrinsic structure and patterns, providing intuitive insights for subsequent model construction. This article first uses Pearson correlation coefficient, judgment matrix, and network graph methods to analyze the correlation between industries and determine the strength of the correlation between different industries. Subsequently, this article uses mixed integer linear programming and branch and bound algorithm to solve the optimal tourism development strategy. Sensitivity analysis shows that ecological investment, government expenditure, and total revenue are the most important factors affecting the performance of the model.

Keywords: Mixed Integer Linear Programming, Optimization Suggestions, Multi-Objective Optimization.

1. Introduction

China's industrial landscape is undergoing a significant shift from traditional manufacturing toward high-tech and service-oriented sectors. Rapid advancements in fields like artificial intelligence, biotechnology, and renewable energy are reshaping both China's economy and the global economic framework. The national economy spans diverse sectors such as agriculture, manufacturing, finance, and real estate, forming a robust and interconnected economic system that highlights balanced national development. The relationships between industries are complex, with the potential to either support mutual progress or create constraints. Investment plays a vital role in job creation, fostering industry growth, boosting economic expansion, and maintaining employment stability. In this study, Lagrange interpolation was initially used to address missing values in the data to ensure the accuracy of the analysis. Subsequently, box plots are used to identify and manage outliers, thereby improving data quality. Then, use scatter plots and pairwise relationship graphs to explore the data, observe its internal structure and patterns, and provide intuitive insights for subsequent model construction. The Pearson coefficient is used to measure the linear, nonlinear, and interdependent relationship between two continuous variables. Use a judgment matrix to determine relevant indicators. Network diagrams are used to visualize the strength of relationships between different industries. Afterwards, support vector machine (SVM) is chosen for model training to analyze data and determine the interrelationships between industries and their impact on economic development.

2. Using polynomial interpolation algorithm to fill in missing values

At the beginning of the data analysis phase, this study conducted a detailed check on the integrity of the dataset. The inspection revealed instances of missing data in the dataset. To address this issue, the study used Lagrange interpolation to fill in the gaps, ensuring the continuity and integrity of the dataset[1]. Lagrange interpolation is a widely used and effective method for handling missing values in discrete data and function approximations. The basic concept when dealing with a set of discrete data points containing missing values is to use a known data point constructor. This function aims to approximate known data as closely as possible and then be used to estimate missing values.

$$P(x) = \sum_{i=0}^{n} y_i l_i(x) \tag{1}$$

where $P(x)$ represents the interpolation at x, y_i is the value of a known data point, and $l_i(x)$ is a Lagrange basis polynomial defined as:

$$l_i(x) = \prod 0 < j < n \frac{x-x_j}{x_i-x_j} \tag{2}$$

where X_i and X_j are the independent variable values of the known data points respectively, and n is the number of known data points minus one.

2.1. Anomaly Detection

Next, this article conducts anomaly detection on a set of data by assigning the upper and lower quartiles as Q_3 and Q_1, respectively. The interquartile range $Q_3 - Q_1$ is represented as q. Outliers are usually defined as points outside the interval, such as $[Q_1 - 1.5q, Q_3 + 1.5q]$ Box plots are particularly effective in representing these critical values, therefore, they are used to detect outliers in the dataset. The box diagram created for outlier detection.

Upon scrutinizing the box plots for each characteristic, it is ascertained that the dispersion of data across all features resides within an acceptable spectrum, with no anomalies identified. Consequently, this study obviates the necessity for outlier scrutiny and remediation, permitting the data to be seamlessly utilized for ensuing analytical procedures [2].

Subsequently, this article analyzed the correlation between different features using scatter plots and used color coding based on different features to illustrate the distribution of data points belonging to different categories in two-dimensional space, as shown in Figure 1.

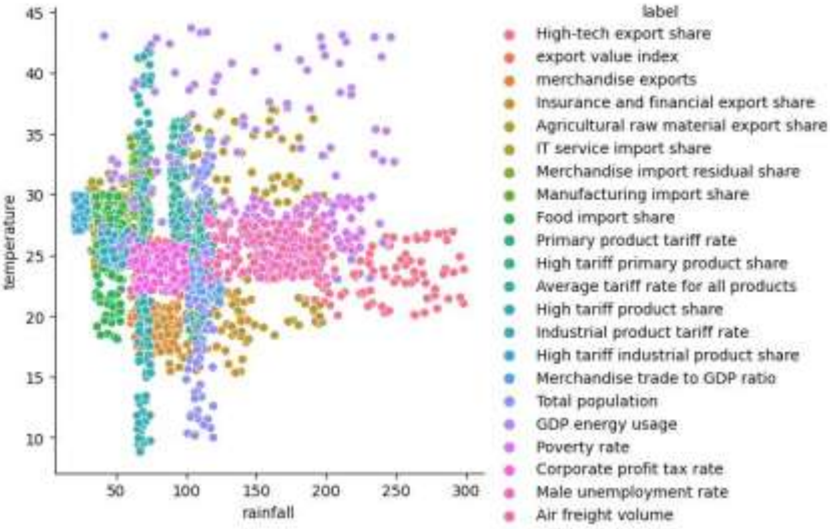

Figure 1. Scatter plot of data distribution

To gain a deeper understanding of the interrelationships and distribution characteristics among the variables in the dataset, this paper employs a pairwise relationship plot for visual analysis. The observations in the dataset are colored and distinguished according to their categories.

The generated pairwise relationship plots reveal that most variables exhibit varying degrees of linear or nonlinear relationships. The exploratory visual analysis of the data significantly enhances this paper's insight into the central tendencies and variability within the dataset, while also providing a clear perspective on the trends of inventory changes. This analytical process is crucial for guiding subsequent data feature extraction, the formulation of analysis strategies, and the precise selection of models.

2.2. Model Establishment and Solution

2.2.1. Solution Thinking

In this study, we initially employed Lagrange interpolation to address missing values in the data, ensuring the accuracy of the analysis. Subsequently, box plots were utilized to identify and manage outliers, thereby enhancing data quality. We then used scatter plots and pairwise relationship plots to explore the data, observing its intrinsic structure and patterns, providing intuitive insights for subsequent model construction. The Pearson coefficient was used to measure the linear, nonlinear, and interdependent relationships between two continuous variables. A judgment matrix was employed to identify relevant indicators. Network diagrams were used to visualize the strength of associations between different industries. Following this, (SVM) were selected for model training to analyze the data and derive the interrelationships between industries and their impacts on economic development.

2.2.2. Correlation Analysis

The examination of correlational aspects involves discerning if there exists a particular interdependence among observed phenomena, delving into the specifics of such dependencies in terms of their orientation and magnitude. This serves as a statistical technique to investigate the associative nature between variables. Utilizing this approach enables researchers to pinpoint and evaluate the distinct influences that various attributes exert on economic progression, thereby furnishing a foundational understanding for devising more impactful methodologies. For the purpose of quantitatively gauging the intensity and vector of linear associations amongst variables, the Pearson correlation coefficient was utilized to conduct an analytical assessment of the linear interconnections present between variables [3].

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{3}$$

where X_i and Y_i are the values for each observed pair.

Through the computation of correlation coefficients for every dyad of variables, this research ascertained the Pearson correlation coefficients for each attribute, thereby signifying the notable correlation existing between the variables and economic advancement. The analytical outcomes reveal that the most pronounced positive correlation is observed amongst high-tech industries, manufacturing, and the service sector, whereas the correlation coefficients for the real estate and chemical industries are comparatively diminutive. These quantitative insights further corroborate the visual observations made in his study and offer a robust statistical foundation for comprehending the determinants of economic growth. Subsequent to this, the pivotal factors influencing economic development were dissected into numerous singular influencing elements, which were then classified into several clusters predicated on their characteristics. Each cluster corresponds to a stratum, thereby constituting diverse hierarchical levels. An evaluation of the comparative significance of these factors was undertaken by juxtaposing them in pairs at each stratum. The importance of these factors was denoted by values ranging from 1 to 9. To streamline this comparative process, they were arranged into a tabular format, recognized as the judgment matrix. Subsequently, the maximum eigenvalue $\lambda_{\max}$ and its corresponding eigenvector of the judgment matrix are calculated using the product method, power method, or square root method. To assess the consistency of the test matrix, the consistency index CI is calculated.

$$CI = \frac{\lambda_{max}-n}{n-1} \tag{4}$$

The CI is compared with the average random consistency index RI. When the calculated result $CR = CI / RI < 0.1$, it can be considered that the calculation result of the judgment matrix has satisfactory consistency; otherwise, it is necessary to readjust the judgment matrix.

Next, assume that the weights of the first-level indicators are $\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_n$ and the weights of the second-level indicator factors are $\alpha_{i1}, \alpha_{i2}, \alpha_3, \cdots, \alpha_{in}$ in The formula for calculating the total weight is [4]:

$$w_i = \sum_{j=1}^{n} \alpha_{ij}\, \alpha_j, i = 1,2, \dots, m \tag{5}$$

We systematically divided the key factors affecting economic development into five levels, assigning them scores of 1, 2, 3, 4, and 5, respectively. If the indicator weight is between two levels, it can be represented as $(0.5,1.5,2.5,3.5,4.5)$.

Based on different levels of evaluation impact, $h = 1,2,\cdots 5$ are used to represent category numbers, which correspond to 5 categories: very small, small, severe, significant, and very significant.

When $h = a$, the gray number is set as $\otimes_1 \in (0, a, 2a)$, and the self-weighting function is defined as [5]:

$$f_1(d_{ij}^k) = \frac{d_{ijh}}{a}, d_{ijh} \in [0, a] \tag{6}$$

$$f_1(d_{ij}^k) = \frac{2a-d_{ijh}}{a}, d_{ijh} \in [a, 2a] \tag{7}$$

The evaluation of different criteria is calculated d_{ijh}, and each scoring result is considered as the evaluation value of the criterion in the $h-th$ gray category function, which can be represented by the following formulas:

$$X_{ijh} = \sum_{x=1}^{p} f_h(d_{ijk}) \tag{8}$$

$$X_{ij} = \sum_{h=1}^{5} X_{ijh} \tag{9}$$

Generally speaking, γ_{ij} represents the sensitivity of the evaluation criterion to the gray level h. The larger the value of γ_{ij}.

The more it can represent the gray level h of the evaluation criterion.

$$R_i = \begin{pmatrix} r_{i1} \\ r_{i2} \\ r_{i3} \\ r_{i4} \\ r_{i5} \end{pmatrix} = \begin{pmatrix} r_{i11} & \cdots & r_{i15} \\ \vdots & \ddots & \vdots \\ r_{ij1} & \cdots & r_{ij5} \end{pmatrix} \tag{10}$$

Subsequently, we utilized heat maps to further illustrate the correlations between features. Heat maps are an effective data visualization tool that uses the intensity of colors to represent the magnitude of values, thereby allowing us to intuitively perceive the strength of correlations between different features.

Upon examining the network diagram and identifying the connections between nodes within the data, it was concluded that the most robust positive correlations are present between GDP, the value added by the secondary industry, and the value added by the tertiary industry. Conversely, the correlation coefficients for other manufacturing sectors, textiles and apparel, and mining industries were found to be comparatively lower.

2.3. Model Establishment

In this study, a variety of algorithms were employed for pre-training to identify the most suitable model for industry correlation analysis. Below, a detailed comparative analysis of the performance of decision trees, support vector machines, linear regression, grey prediction models, and ARIMA models is presented.

Table 1. Model Comparison Table

Model	MAE	MSE	R^2
Linear Regression	0.015792	0.000403	0.744877

According to Table 1, following an extensive application and comparison of a variety of algorithms, this study ultimately determined that the Support Vector Machine (SVM) is the optimal model. Consequently, this paper decides to utilize the SVM to establish a correlation prediction model.

Support Vector Machines (SVM) are robust classification models utilized to address data classification and regression issues. They are highly popular in the realm of machine learning, particularly excelling in handling small sample sizes, non-linear data, and high-dimensional data. When tackling classification problems, the objective of SVM is to identify an optimal decision boundary, namely the hyperplane, that maximizes the margin between different classes, thereby distinguishing them effectively.

Support Vector Machines (SVM) employ kernel functions to map input data into a high-dimensional feature space, thereby seeking the optimal decision boundary within that space.

Table 2. Comparison of Kernel Function Performance

Kernel Function	Score
Linear Kernel	0.963839
Polynomial Kernel	0.963634
Gaussian Kernel	0.974263
Sigmoid Kernel	0.988866

According to Table 2, after comparing various kernel functions, this paper selected the Sigmoid kernel function for the subsequent training and optimization of the model. This choice was based on its demonstrated classification performance on a specific dataset.

$$K(x, z) = tanh(ax^T + c) \tag{11}$$

After selecting the appropriate kernel function, this paper proceeded to construct the model:

$$min \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} a_i y_i = 0 \tag{12}$$

$$f(x) = sign(\sum_{i=1}^{N} a_i^* y_i K(x, x_i) + b^*) \tag{13}$$

Among them, a_i is the Lagrange multiplier, which is used for weight adjustment in the optimization process. y_i represents the label of sample i, N represents the number of samples, and C represents the regularization parameter to control the degree of penalty for model misclassification.

This paper selects the optimal degree of penalty by continuously adjusting the regularization parameter.

Table 3. Comparison of Regularization Parameters

Regularization Parameter	Score
1	0.967950
10	0.968987
100	0.984242
1000	0.981312

According to Table 3, By comparing the performance under different parameter configurations, this paper concludes that on the dataset of this study, the Sigmoid kernel function combined with an appropriate regularization intensity can effectively improve the accuracy of classification[6], as shown in Figure 2.

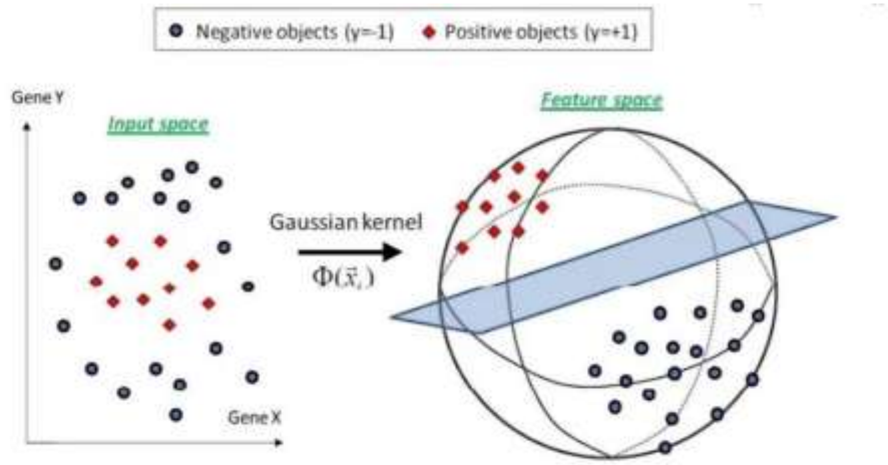

Figure 2. SVM Schematic Diagram

Finally, this article fitted the predicted and actual values, as shown in Figure 3.

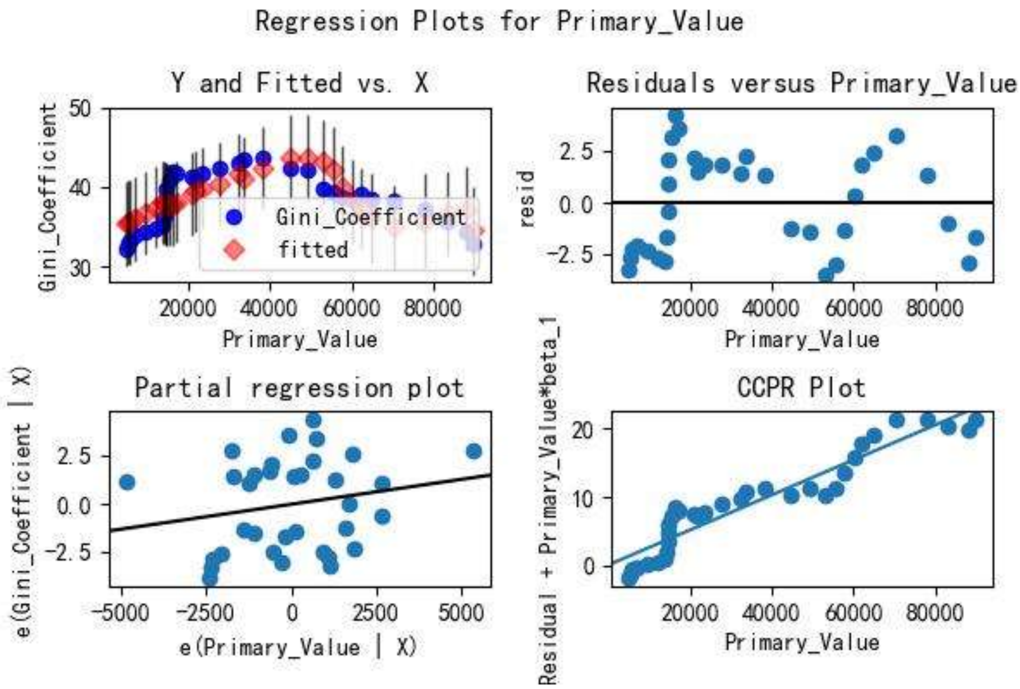

Figure 3. Fitting Effect Graph

3. Conclusions

This study systematically analyzed sustainable development strategies for Juneau's tourism industry by integrating multi-dimensional data and methodologies, including Lagrange interpolation for data imputation, box plots for outlier detection, Pearson correlation analysis for industrial relationships, and support vector machines (SVM) for modeling. Results reveal strong positive correlations among GDP, secondary industry value added, and tertiary industry value added, with high-tech industries and manufacturing mutually driving economic growth, while traditional manufacturing shows negative correlations with high-tech sectors, indicating structural transformation challenges. The optimized SVM model demonstrates robust accuracy in capturing non-linear dynamics, providing a reliable tool for tourism development prediction.

The findings offer practical feasibility for policy formulation: targeted investments in high-tech and service sectors can enhance economic diversification, while ecological investment—identified as a key factor via sensitivity analysis—should be prioritized to balance environmental and economic sustainability. For future research, expanding datasets to include real-time metrics, exploring hybrid machine learning models, and investigating cross-industry spillover effects would further strengthen the applicability of this framework under evolving global dynamics.

References

[1] Xu N, Li Z, Guo N, et al. Global Path Planning for Articulated Steering Tractor Based on Multi-Objective Hybrid Algorithm[J]. Sensors (Basel, Switzerland), 2024, 24(15): 4832-4832.

[2] Chen X, Yuan Z, Feng S. Anomaly detection based on improved k-nearest neighbor rough sets[J]. International Journal of Approximate Reasoning, 2025, 176109323-109323.

[3] Kunekar P, Gupta K M, Gaur P. Detection of epileptic seizure in EEG signals using machine learning and deep learning techniques[J]. Journal of Engineering and Applied Science, 2024, 71(1): 21-21.

[4] Chen X, Suzuki T, Yang H, et al. Phase Transformation and Mechanical Properties of G Phase (Mn6Ni16Si7) in Mn–Ni–Si Model Alloys after 1,000°C Annealing: Microstructure of Materials[J]. MATERIALS TRANSACTIONS, 2024, 65(10): 1234-1238.

[5] Arfa A, Minaei M. Utilizing multitemporal indices and spectral bands of Sentinel-2 to enhance land use and land cover classification with random forest and support vector machine[J]. Advances in Space Research, 2024, 74(11): 5580-5590.

[6] Raza J, Lund A L, Ashraf H, et al. Fuzzy TOPSIS optimization of MHD trihybrid nanofluid in heat pipes[J]. Case Studies in Thermal Engineering, 2024, 64105493-105493.