International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.07
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Analysis of Factors Influencing Wait Times at Disneyland Resort

He Wei

Safety and Emergency Response Institute, Wuhan University of Technology, Wuhan 430070, China

ABSTRACT

This paper aims to enhance the accuracy of wait time predictions and optimize park management at Disneyland, a popular leisure destination facing long wait times for attractions. These wait times adversely affect visitors' immediate experiences and overall evaluations, influencing their willingness to return. These wait times adversely affect visitors' immediate experiences and overall evaluations, influencing their willingness to return. The study employs advanced data analysis and machine learning algorithms to develop high-precision wait time prediction models. Utilizing big data techniques, the research integrates multi-dimensional data, including weather, visitor flow, and historical wait times, to optimize resource allocation and improve operational efficiency. optimize resource allocation and improve operational efficiency and visitor satisfaction. The study further explores the application of these models in other domains such as healthcare, healthcare, and healthcare services. The study further explores the application of these models in other domains such as healthcare and transportation, demonstrating the potential for cross-disciplinary technology transfer to enhance overall service quality and efficiency. The findings underscore the significant impact of big data behavior analysis in theme park management, contributing to the findings underscore the significant impact of big data behavior analysis in theme park management, contributing to increased visitor satisfaction, enhanced competitiveness, and long-term sustainability of the park.

KEYWORDS

Wait time prediction; Big data behavior analysis; Machine learning; Theme park management; Operational efficiency

1. INTRODUCTION

1.1. Background

Disney, as a popular leisure and entertainment destination today, long queues for amusement programs have always been a problem for the park. This not only affects visitors' immediate experience, but also their overall evaluation of the park and their willingness to return. Therefore, how to effectively predict and shorten the waiting time of visitors to enhance visitor satisfaction and park service quality is an issue that deserves in-depth exploration. This study attempts to use advanced data analysis and machine learning algorithms to significantly improve the accuracy of waiting time prediction and optimize park management. With the advancement of science and technology, the application of big data technology in theme park management is promising, which is of great significance to enhance visitor satisfaction and park competitiveness, as well as to ensure the long-term development of the park.

This study attempts to significantly improve the accuracy of waiting time prediction and optimize park management using advanced data analysis and machine learning algorithms. Through big data technology, a high-precision waiting time prediction model is developed to optimize theme park resource allocation and improve operational efficiency and visitor experience. Use machine learning and deep learning algorithms to improve prediction accuracy and apply them to personalized itinerary planning, real-time display systems and virtual queuing systems to reduce waiting time and enhance visitor satisfaction. At the same time, we explore the application of models in medical and transportation fields to promote cross-discipline technology promotion and improve overall service quality and efficiency. With the advancement of science and technology, the application of big data technology in theme park management is promising, providing new solutions to enhance visitor satisfaction and park competitiveness.

1.2. Current Status of the Study

Waiting time prediction has a wide range of applications in the healthcare, transportation, and service industries, where research focuses on how to improve resource allocation efficiency and customer satisfaction by optimizing queuing systems and reducing waiting times.

In China, researchers mainly study the queuing problem of amusement parks through queuing theory and optimization systems. For example, Huang Yuzhuang et al. (2019) established a multi-objective simultaneous queuing optimization model based on the M/M/n queuing theory theoretical model and designed a queuing system based on queuing theory for Android amusement park projects to improve the operational efficiency and visitor satisfaction of amusement parks. Li Zhuo et al. (2021) found the problems of long queuing time and poor experience in amusement parks through investigation, constructed an M/M/1 queuing theory model based on queuing theory, and found that the main problem lies in queuing imbalance through Monte Carlo simulation, and proposed a Web-based queuing system to improve the imbalance in the spatial and temporal distribution of queuing among projects. In addition, Pan Cherry Dan et al. (2019) studied the queuing time prediction algorithm based on BP neural network, which improves the prediction accuracy and optimizes the resource allocation through machine learning method.

Elsewhere, researchers are also advancing research on waiting time prediction. Tasar. B et al. (2020) developed a simulation model for managing waiting time of restaurant customers to optimize the restaurant operation strategy through scenario analysis. Sowndharya. P.K et al. (2024) used M/M/c queuing model to study how to reduce the waiting time of patients in healthcare services by optimizing the server configuration significantly improved service efficiency. In the same year, Isaac. J.V et al. (2024) used a hybrid optimization and discrete event simulation model to reduce waiting times in primary care centers, demonstrating the role of simulation in optimizing healthcare resource allocation.

Despite significant progress in waiting time prediction, there are still research gaps. First, there is a lack of exchange of methods and knowledge in different fields, e.g., prediction methods in healthcare can be applied to transportation, but research is relatively isolated. Second, many studies rely on historical data, but there is insufficient research on model adaptation in real-time data processing and dynamic environments, such as how to use real-time traffic flow data to improve prediction accuracy. Finally, complex models such as deep learning are highly accurate but less explanatory, which is particularly important in areas that require transparency, such as healthcare and banking.

1.3. Innovative Research

1.3.1. Simulation and Optimization Strategies

(1) Designing Simulation Objects and Processes

Determine visitor behavior patterns, such as arrival time, queuing time, ride selection, etc. Design ride operation processes, including opening hours, capacity limits, and maintenance schedules. Consider the impact of weather changes on visitor behavior and facility operations, such as bad weather leading to fewer visitors and facility closures. These designs simulate real-life scenarios and provide data support for optimization strategies.

(2) Simulation Scenario Design

Demonstrate operations under different conditions, such as peak periods, different weather, weekdays vs. weekends. The peak period simulates a large influx of tourists to analyze queuing and waiting time; different weather conditions analyze the impact of weather on queuing; and weekdays and weekends analyze the difference in traffic flow and queuing changes. These scenarios help analyze waiting times and provide data support for optimization strategies.

1.3.2. Optimization Strategy Recommendations

Based on the simulation results, optimization strategies are proposed. Increase service staff during peak periods and bad weather to reduce queuing time; reduce the opening of facilities when there are fewer visitors to reduce costs; and optimize attraction opening times to balance visitor distribution and reduce crowding. These strategies enhance amusement park operational efficiency and visitor satisfaction.

1.3.3. Attempts at forecasting models such as multidimensional time series

Explore integrated models (e.g., LSTM) that combine weather, historical data, and special events to make predictions. Multidimensional time series models consider multiple factors to improve prediction accuracy and stability. Combine weather, historical queuing, and special event data to build an integrated prediction model to improve performance. Predict waiting time more accurately to provide better trip planning suggestions for tourists and improve the overall experience.

2. EXPERIMENTAL TREATMENT

2.1. Data Sources

The dataset used in this study was obtained from the Kaggle website and contains data on "Expected Wait Times for Popular Attractions at Walt Disney World", which is designed to analyze and predict wait times for popular attractions at the four theme parks at Walt Disney World (WDW) in Orlando, Florida. The dataset covers the years 2012 through 2022, with 10 years of detailed wait time information collected from "touringplans.com". Analyzing this data reveals the changing patterns of wait times at each attraction over time and under different conditions, which can support the improvement of theme park operational efficiency and visitor experience.

The dataset contains wait time records for 13 rides over 17 seasons, yielding 884 unique combinations for analysis. Each combination not only includes wait times, but also integrates other information that may affect wait times, such as daily operational data, event schedules, and weather conditions. This combined data helps to accurately analyze the impact of various factors on wait times.

2.2. Data Pre-processing

Data preprocessing is an indispensable part of the process of constructing waiting time prediction models. The accuracy and stability of the model can be significantly improved through steps such as data cleaning, format conversion, feature extraction and coding. Due to the huge amount of data involved in this study, it is difficult to be processed efficiently in a stand-alone Python environment, so we adopted big data techniques for data processing. Data preprocessing was accomplished in Spark's Jupyter Notebook environment with the following main steps:

First, we read multiple datasets that record the attendance of visitors to Disneyland, daily event schedules, etc. These multi-dimensional data provide comprehensive information to support the construction of the prediction model.

Table 1. Introduction to all tables

File Name	Description
attendance.csv	Records the daily performance and parade times, including the start and end times for two parades.
entity_schedule.csv	Contains information about the opening and closing times of parks and attractions, as well as updates. Columns include entity name, type, start time, end time, update time, and work date.
glossary.xlsx	Provides detailed explanations of data fields, including their names and meanings. This file helps understand and reference other datasets.
link_attraction_park.csv	Records the correspondence between parks and attractions, including attraction names and the parks they belong to.
parade_night_show.csv	Similar to attendance.csv, records daily performance and parade times, including the start and end times for two parades.
weather_data.csv	Contains daily weather data, such as temperature, humidity, wind speed. Helps analyze the impact of weather on attendance and waiting times.
waiting_times.csv	Records waiting times for various attractions, including queue length, unit time, load, capacity, adjustment times, start time, run time, stop time, and the maximum capacity of the day.

We harmonized the temporal format of all datasets to ensure temporal consistency and facilitate subsequent data integration and analysis. Subsequently, data cleansing was performed to remove duplicate values and deal with missing values to ensure data continuity and integrity. On this basis, key temporal features, such as the day of the week, whether it is a weekend, month and quarter, were extracted to capture cyclical variations in waiting times, as well as holiday features, as holidays have a significant impact on the number of visitors and waiting times. Data sets from different sources are merged to form a comprehensive dataset containing weather, waiting times, daily activity schedules, rides, and park information to ensure that each record has comprehensive feature information. Numerical features are normalized to reduce the impact of different feature scales on model training, and category features are converted to numerical form by solo thermal coding to enable the model to make full use of this information. Finally, a comprehensive dataset after cleaning, feature extraction and coding are obtained, which lays a solid foundation for constructing a high-precision waiting time prediction model.

3. MODELING AND FORECASTING

3.1. Weather Factors and Waiting Time Modeling

3.1.1. Model Training and Evaluation

In the model training phase, we selected models such as linear regression and random forest regression, and performed hyperparametric grid search to optimize the performance of the model. The training set and test set of the model are divided in the ratio of 80% and 20%. After training and evaluation, the more typical models are selected, and the performance is as follows:

Table 2. Model training content

Model Type	Training Set RMSE	Training Set MAE	Test Set RMSE	Test Set MAE
Linear Regression	12.48	8.53	12.48	8.53
Random Forest Model	12.25	7.97	12.25	7.97

3.1.2. Analysis of experimental results

By analyzing the training results, we found that weather factors have a more significant effect on waiting time in Disneyland. Specifically, meteorological variables such as temperature and humidity affected visitors' willingness to travel and the waiting time for each attraction to a certain extent. In addition, time characteristics (e.g., weekends and holidays) also have a significant impact on waiting time, which is generally longer on weekends and holidays.

Table 3. Model training results

Model Type	Dataset	RMSE	MAE
Linear Regression	Training Set	12.48	8.53
Linear Regression	Test Set	12.48	8.53
Random Forest Model	Training Set	12.25	7.97
Random Forest Model	Test Set	12.25	7.97

These metrics indicate that the random forest model outperforms the linear regression model on both the training and test sets and is able to more accurately predict wait times at Disneyland.

3.1.3. Predictive condition design

In summary, we design some samples to test and prioritize the use of random forest regression model for waiting time prediction.

3.2. Designing Samples for Different Weather and Time Conditions

In order to fully evaluate the performance of the model in various situations, we designed different samples representing common weather and time conditions.

Table 4. Sample condition

Sample Conditions	Temperature (°C)	Humidity (%)	Time	Season
Normal Conditions	25	50	Wed	Spring
High Temperature Low Humidity, Fri	35	30	Fri	Summer
Low Temperature High Humidity, Mon	10	80	Mon	Fall
Winter Low Temperature, Sat	5	60	Sat	Winter
Fall High Temperature High Humidity, Thu	30	70	Thu	Fall
Mild Weather, Wed	20	50	Wed	Spring
Extreme Low Temperature High Humidity, Thu	-5	90	Thu	Winter

3.2.1. Analysis of projected results

By analyzing the sample prediction results, we found the following:

Table 5. Sample prediction results

Sample Conditions	Predicted Waiting Time (minutes)
Normal Conditions	7.36
High Temperature Low Humidity, Fri	7.36
Low Temperature High Humidity, Mon	7.36
Winter Low Temperature, Sat	6.54
Fall High Temperature High Humidity, Thu	7.36
Mild Weather, Wed	7.36
Extreme Low Temperature High Humidity, Thu	5.54

The above analysis shows that weather and time conditions do have a significant impact on wait times at Disneyland. In particular, wait times decreased significantly under extreme weather conditions, such as extreme low temperatures and high humidity, which may be related to the reduced willingness of visitors to travel under severe weather conditions. Changes under other conditions were less pronounced than expected, suggesting that visitor travel behavior and wait times are relatively stable under regular weather and time conditions.

By further optimizing the model and adding more samples, the accuracy and stability of the prediction can be improved, providing more reliable data support and decision-making basis for the operation and management of Disneyland.

3.3. Operational Factors and Waiting Time Modeling

3.3.1. Model Training and Evaluation

In the model training phase, we selected linear regression, decision tree, random forest and neural network models with hyperparameter tuning to optimize the performance of the models. The model's training set and test set were divided in the ratio of 80% and 20%.

3.3.2. Analysis of experimental results

By analyzing the training results, we found that different operational factors have a more significant impact on the waiting time in Disneyland. Specifically, operational variables such as different types of attractions, operating hours, and maintenance hours affected visitors' waiting times to some extent. In addition, time characteristics (e.g., weekends and holidays) also have a significant impact on waiting times, with longer waiting times prevailing on weekends and holidays.

3.3.3. Model evaluation indicators

In order to quantify the performance of the model, we used root mean square error (RMSE) and mean absolute error (MAE) as evaluation metrics. The specific results are as follows:

Table 6. Model training results

Model Type	Training Set RMSE	Training Set MAE	Test Set RMSE	Test Set MAE
Linear Regression	19.82	13.12	19.79	13.09
Decision Tree Model	7.43	2.28	13.30	5.63
Random Forest Model	17.36	10.49	17.33	10.46
Neural Network Model	15.87	9.89	15.83	9.86

These metrics indicate that the Random Forest model performs relatively well on both the training and test sets and is able to predict wait times at Disneyland more accurately.

3.3.4. Analysis of projected results

In order to evaluate the performance of the model under different conditions, we designed a number of samples for testing. The sample conditions include different facility types, operating hours, maintenance hours, etc.

Table 7. Sample condition

Sample Conditions	Facility Type	Operation Time	Maintenance Time	Day	Season
Facility A, Normal Operation Time, No Maintenance	A	Normal	None	Wed	Spring
Facility B, Peak Operation Time, No Maintenance	B	Peak	None	Fri	Summer
Facility C, Low Operation Time, With Maintenance	C	Low	Yes	Mon	Fall
Facility D, Normal Operation Time, With Maintenance	D	Normal	Yes	Sat	Winter
Facility E, Peak Operation Time, With Maintenance	E	Peak	Yes	Thu	Fall

Table 8. Sample prediction results

Model Type	Sample 1 (minutes)	Sample 2 (minutes)	Sample 3 (minutes)	Sample 4 (minutes)	Sample 5 (minutes)
Linear Regression	13.31	12.65	16.08	9.79	13.04
Decision Tree Model	25.00	0.00	5.00	25.00	25.00
Random Forest Model	8.36	5.74	5.74	8.36	8.36
Neural Network Model	5.13	13.56	13.37	9.62	12.87

The above analysis shows that different operational factors have a significant impact on wait times at Disneyland. In particular, factors such as facility type, hours of operation, and hours of maintenance significantly influence predicted wait times. In addition, temporal characteristics (e.g., weekends and holidays) also have a significant impact on wait times. Different models vary in their ability to capture these features, and the random forest model performs relatively well on the training and test sets with good generalization ability.

By further optimizing the model and adding more samples, the accuracy and stability of the prediction can be improved, providing more reliable data support and decision-making basis for the operation and management of Disneyland.

3.4. Impact of Visitor Numbers on Waiting Times

3.4.1. Model Training and Evaluation

In this section, we have trained and evaluated the data using linear regression, decision tree, random forest and neural network models. The dataset is divided into training set (80%) and testing set (20%). The training and testing performance of the models is shown below:

Table 9. Model training results

Model Type	Training Set RMSE	Training Set MAE	Test Set RMSE	Test Set MAE
Linear Regression	18.32	12.49	18.34	12.49
Decision Tree Model	4.63	1.10	13.66	6.34
Random Forest Model	16.77	10.09	16.78	10.09
Neural Network Model	15.90	9.61	15.92	9.62

3.4.2. Analysis of experimental results

By comparing the performance of different models, the following points can be observed:

The linear regression model performed relatively poorly, with negative values in the prediction results, indicating that the model failed to fit the data well. The decision tree model performs very well on the training set, with possible overfitting. The error on the test set is significantly higher. The random forest model has a more balanced performance on the training and test sets with better generalization. The performance of the neural network model is also more balanced and shows better prediction ability with relatively low errors on both the training and test sets.

3.4.3. Model evaluation indicators

In order to more fully evaluate the performance of the model under different conditions, we designed the following five samples representing different visitor numbers and facility operating conditions.

Table 10. Sample condition (Time unit: minutes)

Sample	Visitor Count	Facility Operation Units	Carrying Capacity	Total Capacity	Adjusted Capacity	Open Time (min)	Operating Time (min)	Downtime (min)	Maximum Unit Count	Workday
1	50,000	5	500	1,000	900	60	30	5	10	Monday
2	80,000	10	800	1,500	1,400	60	30	5	20	Friday
3	60,000	7	600	1,200	1,100	60	30	5	15	Tuesday
4	45,000	3	450	900	800	60	30	5	10	Saturday
5	100,000	15	1,000	2,000	1,900	60	30	5	25	Thursday

3.4.4. Analysis of projected results

Table 11. Sample prediction results

Model	Sample 1 (minutes)	Sample 2 (minutes)	Sample 3 (minutes)	Sample 4 (minutes)	Sample 5 (minutes)
Linear Regression Model	Not reasonable	Not reasonable	Not reasonable	Not reasonable	Not reasonable
Decision Tree Model	25.00	5.00	5.00	40.00	5.00
Random Forest Model	33.36	23.10	33.36	33.36	23.10
Neural Network Model	615.86	557.18	593.97	624.83	490.27

The above analysis shows that the prediction results of the random forest model under different conditions are relatively more stable and reasonable, and it is the best performing model in this experiment. The neural network model, although with lower errors on training and testing, showed larger predicted values in the sample prediction, which needs to be further optimized and adjusted.

In this study, it is recommended that the random forest model be used preferentially in practical applications for the prediction of waiting time in Disneyland, and the model parameters should be further adjusted and optimized with the actual data in order to enhance the prediction accuracy and stability.

3.5. Impact of Parades and Night Shows on Waiting Times

3.5.1. Model Training and Evaluation

In this section, we have trained and evaluated the data using linear regression, decision tree, random forest and neural network models. The dataset is divided into training set (80%) and testing set (20%). The training and testing performance of the models is shown below:

Table 12. Model training results

Model Type	Training Set RMSE	Training Set MAE	Test Set RMSE	Test Set MAE
Linear Regression	15.98	11.25	16.03	11.26
Decision Tree Model	7.67	3.02	14.08	7.16
Random Forest Model	13.86	8.09	14.06	8.15
Neural Network Model	11.40	6.21	14.06	8.15

3.5.2. Analysis of experimental results

By comparing the performance of different models, the following points can be observed:

The linear regression model performed relatively poorly, with negative values in the prediction results, indicating that the model failed to fit the data well. The decision tree model performs very well on the training set, but may suffer from overfitting. The error on the test set is relatively high. The random forest model has a more balanced performance on the training and test sets, with better generalization ability and more stable prediction results. The performance of the neural network model is also more balanced and shows better prediction ability with relatively low errors on both the training and test sets.

3.5.3. Sample Predictive Analytics

In order to more fully assess the performance of the model under different conditions, we designed the following five samples representing different parade and night performance scheduling conditions.

Table 13. Sample condition

Sample	Operating Unit	Load	Total Capacity	Adjusted Capacity	Opening Time	Operating Time	Downtime	Maximum Unit	Workday	Night Show Time	Tour 1 Time	Tour 2 Time
1	10	500	1,000	950	60	55	5	5	Monday	20:00	17:30	12:10
2	15	700	1,500	1,400	70	65	5	10	Tuesday	21:00	18:00	12:30
3	20	900	2,000	1,950	80	75	5	15	Wednesday	22:00	18:30	13:00
4	25	1,100	2,500	2,400	90	85	5	20	Thursday	21:00	18:00	12:10
5	30	1,300	3,000	2,950	100	95	5	25	Friday	21:00	18:00	12:30

3.5.4. Analysis of projected results

Table 14. Sample prediction results

Sample	Linear Regression	Decision Tree	Random Forest	Neural Network
1	Not reasonable	5.00	5.20	6.01
2	Not reasonable	5.00	5.35	24.43
3	Not reasonable	10.00	6.95	7.92
4	Not reasonable	15.00	5.24	11.12
5	Not reasonable	0.00	19.08	12.50

The above analysis shows that different factors have a significant effect on the waiting time. The higher the number of operating units, the waiting time is relatively shorter as more visitors can be accommodated at the same time. Higher carrying capacity and total capacity help to reduce waiting time but need to ensure efficient operation of the facility. Longer opening hours and operating hours can spread out the flow of visitors and thus reduce waiting time. The scheduling of nighttime performances and parades can affect the distribution of visitors, which indirectly affects the waiting time. Therefore, this study suggests giving priority to the use of the random forest model for the prediction of waiting time in Disneyland in practical applications, and further adjusting and optimizing the model parameters with the actual data in order to improve the prediction accuracy and stability. At the same time, the reasonable arrangement of the operation time of the facilities and the time of the performance activities can effectively reduce the waiting time of the tourists and improve the satisfaction of the tourists.

3.6. Processing and Modeling of LSTM Models

3.6.1. Data preparation and cleansing

Similar to the random forest regression model, the data preparation and cleaning process for the LSTM model also involves waiting for the reading of time data and weather data, date format processing, data merging and feature extraction. In this step, we convert the data into a format suitable for time series analysis so that the LSTM model can capture patterns and trends in the time series.

3.6.2. Feature Engineering and Data Processing

In the feature engineering phase, we normalized the numerical features and coded the category features with unique heat. Subsequently, we converted the dataset to a 3D tensor format (number of samples, number of time steps, number of features) suitable for LSTM modeling.

3.6.3. Model Architecture and Training

The architecture of the LSTM model includes two LSTM layers and one fully connected layer. During the model training process, we set the number of time steps (TIME_STEPS) to 10, the batch size (batch_size) to 32, and performed 10 training rounds (epochs). The training and evaluation results of the model are as follows:

The root mean square error (RMSE) on the training set is 21.02 and the mean absolute error (MAE) is 15.76; The root mean square error (RMSE) on the test set was 21.07 and the mean absolute error (MAE) was 15.79.

3.6.4. Analysis of experimental results

The training and testing results of the LSTM model showed that it did not perform as well as expected in processing time series data. Although the LSTM model is theoretically capable of capturing complex patterns in time series, in this study, the model has a large error and further optimization of the model architecture and parameter settings is needed. In addition, the high dimensionality and complexity of the data may have had an impact on the training of the LSTM model.

By comparing the performance of linear regression, random forest regression model and LSTM model, it can be seen that the random forest regression model has better accuracy and stability in predicting the waiting time of Disneyland. The LSTM model, although it has theoretical advantages in dealing with time-series data, does not perform as well as expected in this study, and may need to be further adjusted and optimized. Taken together, this study suggests prioritizing the use of the random forest regression model for waiting time prediction in practical applications.

3.6.5. Conclusion of the experiment

Through a comparative analysis of multiple models, this study found that the Random Forest model performed best in predicting wait times at Disneyland. Weather factors, operational factors, number of visitors, and parades and nighttime shows all had significant effects on wait times. Specifically, temperature and humidity significantly affected wait times, with changes especially pronounced during extreme weather conditions. Operational factors such as facility type, hours of operation and maintenance also have a significant impact on wait times, and proper operational scheduling can effectively reduce wait times. Waiting time varies significantly between peak and trough periods in terms of the number of visitors, and the pressure during peak periods can be alleviated by appropriately adjusting the number of units operated and the capacity of the facilities. Reasonable scheduling of parades and night performances can spread out the visitor flow and reduce the waiting time at popular facilities. The combination of these factors affects the overall visitor experience, and optimizing these aspects can significantly improve the park's operational efficiency and visitor satisfaction.

Although the LSTM model has the theoretical advantage of dealing with time series data, the application in this study is not satisfactory and needs further adjustment and optimization. Future research can continue to optimize the random forest model and explore other potential models to improve the accuracy and stability of waiting time prediction and provide more reliable data support and decision-making basis for the operation and management of Disneyland.

4. RECOMMENDATIONS AND IMPLICATIONS

In order to enhance the visitor experience and operational efficiency of Disneyland, this paper puts forward the following comprehensive recommendations covering data integration, model optimization, resource scheduling, and customer experience optimization.

First, multi-dimensional data such as weather, visitor flow, and historical waiting time need to be fully integrated to ensure the comprehensiveness and accuracy of the data. The quality of model input is improved through data cleaning, feature extraction and other operations. Data visualization and exploratory data analysis (EDA) are recommended to identify potential patterns and trends.

In terms of model selection and optimization, the best performing random forest model is preferred and optimized, while other models such as LSTM and XGBoost are explored. Regularly evaluate the model performance and optimize the parameters using cross-validation and grid search to improve the generalization ability and prediction accuracy of the model. Use explanatory techniques such as SHAP values and LIME to increase model interpretability.

Resource scheduling and staff allocation optimization are also critical. Based on the forecast results, formulate resource scheduling strategies for peak hours, increase staff or adjust queuing areas. Establish a real-time monitoring system to dynamically adjust resource allocation and staff scheduling, and introduce an intelligent queuing system to provide real-time queuing information through cell phone applications to spread out the queuing pressure during peak hours.

In terms of customer experience optimization and digital marketing, the forecast results can be used to provide personalized recommendations and information pushes to help plan tour itineraries. Enhance visitor interaction by releasing exclusive content and interactive activities through social media and official apps, and organize regular themed events to attract more visitors. Establish a customer feedback mechanism and collect satisfaction survey data to further optimize management strategies.

Continuous improvement is as important as popularization and application. Establish a continuous improvement mechanism to regularly evaluate and optimize forecasting models and strategies. Strengthen cross-departmental collaboration and data sharing to enhance overall operational efficiency. Promote the application of waiting time prediction models to other areas, such as traffic flow prediction and healthcare resource scheduling, to leverage wider practical value and economic benefits. Enhance staff's data analysis and model prediction skills through education and training to ensure effective implementation of strategies.

With these recommendations, the Disney Company can better manage wait times, improve guest experience and operational efficiency, and realize higher customer satisfaction and economic benefits. Not only that, this waiting time prediction model can be modified and applied to other areas, such as healthcare and transportation. Hospitals can use it to better arrange the working hours of doctors and nurses to reduce the waiting time of patients; public transportation departments can also optimize the scheduling and improve the efficiency of the system. Therefore, the study of theme park waiting time prediction not only has theoretical value, but also has a wide range of practical application prospects.

REFERENCES

[1] PAN Cherry Dan, QIAN Jia Li, HE Yan Lei, et al. Research on queuing time prediction algorithm based on BP neural network [J]. Journal of Wenzhou University (Natural Science Edition), 2019, 40(03):28-35.

[2] S. Zhao. Research on bus schedule optimization and on-time control based on arrival time prediction [D]. South China University of Technology, 2022. DOI:10.27151/d.cnki.ghnlu.2022.004941.

[3] GUO Linlin, WEI Yongxiang, GUO Linying,et al. A machine learning-based study of waiting time prediction and correlation analysis for pediatric emergency room patients [J]. Journal of Medical Informatics, 2022, 43(04):33-39.

[4] Annie Wei, Ning Zhao, Zhijian Zhang. Prediction of waiting time for tandem queuing system based on machine learning[J]. Journal of Southwest Normal University (Natural Science Edition), 2022, 47(12):11-21. DOI:10.13718/j.cnki.xsxb.2022.12.002.

[5] Tasar B, Ventura K, Cicekli G U. A simulation model for managing customer waiting time in restaurants: scenarios and beyond [J]. 2020, 122(9):2881-2894.

[6] Sowndharya P K, Bagyam A E J. Optimal server analysis of M/M/c queueing model to reduce the waiting time of patients in healthcare service [J]. International Journal of Mathematics in Operational Research, 2024, 27(2):254-266.

[7] Isaac J V, Eduardo L C, C. J V, et al. Hybrid optimization and discrete-event simulation model to reduce waiting times in a primary health center [J]. Expert Systems With Applications, 2024, 238(PB).