IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105346 |

Tourist Prediction Using Machine Learning Algorithms

Y.Monica, Vijay Sareadhi

Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad (TS), India

ABSTRACT: The purpose of the project is to forecast tourism using machine learning techniques. Tourism significantly improves travellers' ability to understand and experience the culture, language, customs, and way of life of the people who live in the chosen nation. Technology innovation is influencing all industries, including tourism, and paving the road for machine learning. The goal of this system is to promote cross-cultural tourism by enhancing the visitor experience and fostering cultural pride. Increasing understanding of other cultures, traditional knowledge, practises, and lifestyles are the goals of this strategy, which promotes fruitful interactions between visitors and locals. To do this, we forecasted the expected amount of tourists visiting India implementing algorithms that belong to the Naive Bayesian, Random Forest, and Decision Tree classes of machine learning. We applied these after researching previous research on the application of machine learning in the tourism business, we developed a set of techniques and used them to a dataset of 48 rows and 42 columns. Our results demonstrated good accuracy when predicting the total amount of tourists using the dataset.

I.INTRODUCTION

The significance of tourism in the world aids travellers in acclimating to the customs, language, and way of life of the locals at their location. The benefits of tourism include the generation of jobs, profits in foreign currencies, improvement of infrastructure, eradication of poverty, lowering of inequality, and regional development which is balanced. Furthermore, encouragement of international peace is dependent upon tourism. The tourism and travel industry has been significantly impacted by the technological innovation that machine learning is fostering across all industries. It is not difficult to fundamentally modify commercialization methods and how the travel industry functions. Forecasting for the tourist industry has recently caught the interest of many scholars because tourism is crucial to national economies. Regression and time-series analysis methods mostly make use of prediction approaches that were common in prior research procedures. Modern approaches like machine learning can significantly advance this field, even though these traditional methods have showed some promise in tourism forecasting. In reality, a number of forecasting applications, like tourism forecasting, have successfully incorporated machine learning techniques. Section II of the paper includes literature on machine learning and the tourism industry. Section III discusses machine learning theory, particularly for prediction. One of them is the use of machine learning in the travel sector. After contrasting machine learning strategies for predicting tourists, a conclusion is offered.

II.LITERATURE REVIEW

- In their study, Tasfiqul Ghani et al. (2018) employed a smart phone application to integrate machine learning into its programe, It discovers every location of interest the database's administrator has added. This software allows users to sign in with either their Facebook or basic information. They can uncheck the attractions they do not want to see. The application's machine learning algorithms will determine the most efficient paths to each destination individually. Users can also use their phones' offline databases to compose notes that will be kept as entries in their diaries. Users can choose to make those entries publicly accessible or to prevent others from sharing them.
- Indri Hapsari et al. (2018) created the visiting time prediction model as part of their research. Because there are six parameters that affect the length of time spent visiting, Multiple linear regression evolves into a popular model when considering factors things like access, government, rating, the number of reviews, the number of photographs, and other information. These elements serve as the independent variables that forecast the dependent ones, such as visiting time. Additionally, by utilising Expectation Maximisation to convert they used Ordinal Logistic Regression (OLR), a technique for converting interval data into ordinal data. Then they applied five of the most popular machine

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105346 |

learning algorithms, adding Multi-Layer Perceptrons, the use of Support Vector Machines, Decision Trees, Linear Regression, and k-Nearest neighbours to the classification process.

- Three methods to forecast journey time were proposed by the study they conducted, Da-Jie Lin et al. Some examples of these approaches include Linear Regression (LR), Gradient Boosting Regression Trees (GBRT), and K-Nearest Neighbours (KNN). The findings indicated that every machine learning technique utilised in this paper performs well in terms of predicting journey time.
- Mehrbakhsh Nilash et al. (2017) suggested applying collaborative filtering (CF) with several criteria recommendation technique that the travel and tourist sector. They utilized clustering, dimensionality reduction (DR), and prediction approaches, including ANFIS, SVR, PCA, EM, and SOM. By utilising the results of clustering approaches, they also employed HGPA to enhance the multi-criteria CF's accuracy of recommendations.
- In their research, Kyoungho Son et al. (2018) used the LSTM machine learning technology to forecast tourists in a particular tourist destination. LSTM is one of the RNN network variants. favourable outcomes demonstrating its application in a practical setting were obtained.
- In their study, NesreenKamel et al. (2018) investigated multiple machine learning models to address the issue of tourism forecasting and demonstrate the effectiveness of seven widely used machine learning techniques.

III.MACHINE LEARNING ALGORITHMS, PARTICULARLY FOR PREDICTION

The many kinds of machine learning techniques are covered in this part, along with examples of how they have been used to the analysis of tourism-related data. The two machine learning techniques that are utilized most frequently in the travel and tourism sector are Association learning and categorization learning are both typical. In association learning, the learning strategy looks connections or links between different aspects of visitor behaviour. That is often referred to as unsupervised learning. Learning by classification is the second type of machine learning. This learning strategy uses a set of previously classified instances to figure out how to categorise previously unidentified ones. guided learning is what this is. The numerous machine learning methods used in tourism are described in the section that follows. In the tourism industry, machine learning techniques are employed for three things: forecasting travel costs, studying traveller characteristics, and estimating the arrival of tourists are all included. Brief descriptions of the ten machine learning approaches utilised to support these operations are provided in this section.

1. **Logistic Regression** : Using an equation, a huge amount of data is prepared for categorization in logistic regression it is used to forecast discrete values ("Binary integers like 0/1, yes/no, true/false") based on a number of separate factors. The output numbers between 0 and 1 match the expected values thanks to the likelihood forecast. The coefficients produced by all varieties of logistic regression can be used to convert the probability into a log it format.
2. **Linear Regression:** A model (equation) based on data is used in linear regression. Make predictions about one variable using a linear regression model based on certain values of the other variable. That dependent variable, often known as (y), the response variable, is the variable for which predictions are being made. The (independent variable), often known as (x), the predictor variable, is the one that is used to create these predictions.
3. **Decision Tree:** The decision tree is one of the techniques for supervised learning. Regression and classification can both be done using the decision tree. In a decision tree building process, after the root has been assigned to the dataset's best attribute, the training dataset is split into subsets. Data splitting is based on dataset characteristics. This process continues until all data is categorised and a leaf node is located at each branch. It is possible to calculate information gain to determine which characteristic is providing the most information. In order to forecast a class or the value of the target variable, a training model that uses decision trees must be created.
4. **Support vector machine**: The support vector machine algorithm, a binary classifier. In 1995, Vapnik developed the Support Vector Machine methodology. SVM is a machine learning technique that is so well-liked that it may be considered a group unto itself. To establish the limits of a choice among a collection of data points with various label classifications, a separating hyperplane is used. This approach for classification is strictly supervised. To put it another way, the technique uses input or training data to optimise a hyper plane, and this decision-making procedure then categorises fresh cases. Depending on the kernel being utilised, SVM can classify data in both linear and nonlinear ways.
5. **Naive-Bayes:** Bayes theorem-based classification approaches are used to build classifiers using a supervised method known as Naive-Bayes. The naïve Bayesian model is easy to build and especially helpful for large data sets. In this strategy, each feature happens independently of every other feature. Due to the fact that classification only requires a

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105346 |

limited amount of training data and that all terms may be pre computed, it is simple, quick, and effective. Naive Bayes is yet another intricate categorization technique. The posterior probability P(A|B) can be determined using the Bayes theorem from P(A), P(B), and P(B|A). comparable to the following formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

A, B = events
$P(A|B)$ = probability of A given B is true
$P(B|A)$ = probability of B given A is true
$P(A), P(B)$ = the independent probabilities of A and B

6. **KNN:** Regression and classification are handled by this algorithm. In machine learning, it is the most fundamental algorithm. The great majority of its closest k neighbors are checked, and it acquires cases and new data. KNN directly uses the training dataset to create predictions. KNN is referred to be non-parametric since it makes no assumptions about the underlying data distribution.
7. **K-Means Clustering:** is an unsupervised technique that divides items into K numbers of objects based on their qualities and features. K is an integer with a positive sign. The datasets are initially partitioned using Euclidean distance in order to organise them into clusters. Data and the centroid of the relevant cluster's sum of squares are reduced before grouping. In order to categorise the data, K-mean clustering is used.
8. **Random Forest:** This supervised classification method uses random forest. When many decision trees are combined, a random forest algorithm—or collection of different categorization trees is created. Both classification and regression can be done with it. There are certain rule-based systems in every decision tree. For the training dataset with supplied targets and features, the decision tree method contains a set of rules. In contrast to the decision tree, the root node can be found using the random forest technique without the requirement to calculate information gain. To forecast the outcome and save the projected outcome, it applies the rules of every decision tree constructed at random. Additionally, it determines the vote for every expected target. As a result, the random forest's highest-voted prediction is used as the final one.
9. **Algorithms for Dimensionality Reduction:** To reduce the number of random variables, these algorithms employ some significant factors. Techniques for dimensionality reduction include feature extraction and feature selection, to name a couple. A method for separating significant factors from a large number of variables is principal component analysis (PCA). From high dimensional data, it extracts the set of features with low dimension. When we have more than three-dimensional data, it is mostly employed.
10. **Ada Boost and Gradient Boosting Algorithms**: enhancing the gradient and AdaBoost are applied as approaches for regression and classification. Gradient boosting iteratively strengthens the algorithm's foundation, much like decision trees, while AdaBoost chooses features that increase the model's ability for prediction. The three different gradient boosting methods utilised to increase GBM, XGBoost, LightGBM, and Catboost are the predictive models with the highest accuracy.

IV.MACHINE LEARNING APPLICATIONS IN THE TRAVEL INDUSTRY

The services that make up the tourist sector include travel, lodging, and other related services. Every day, billions of dollars are exchanged inside this ecosystem. Its everyday operations rely on technology. There is a constant need for new, cutting-edge technologies in this enormous industry to improve corporate operations and boost profitability. It is a sector that prioritises the wants of the consumer over their necessities. Competitors work to provide their clients with better services at a reduced cost as a result. Machine learning can be used to enhance these tactics.

1. Prediction of seasonal demands for services

When we discuss tourism, we are referring to an industry where demand is cyclical. This temporality may or may not be related to climate seasons. In case, it is at the times when vendors of items associated to tourism have the chance to boost their revenues and are eager to take advantage of this chance. Algorithms for machine learning can be utilised to do it. By examining raw data from the past and making predictions about the future, a computer quickly and reliably determines

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105346 |

the relationship between the components of this seasonal demand. By using historical patterns and machine learning algorithms, a technique known as predictive analytics can forecast future events.

2. Pricing strategies

Competitive pricing is one of the main methods used by sellers of tourism-related goods to draw clients. To draw in as many customers as possible, businesses attempt to offer price modifications without sacrificing their revenues. Machine learning is helpful in this situation. Seasonality, the hotel's or tourist service's past, nearby activities, rivalry or third parties' promotions... Predictive models can be used to analyse all of these data and deliver the most competitive prices, giving businesses a competitive edge.

3. Personalized recommendations

Important travel websites like TripAdvisor and Expedia started using recommendation engines ten years ago to show consumers the vacation packages that best suit their consumer profiles. To provide you with tailored travel recommendations, the engines gather information about your tastes, finances, and personal information. By evaluating possibilities using Machine Learning algorithms, information from diverse sources and service providers is leveraged to locate viable alternatives.

4.Customer experience

The travel industry's clients are diverse, and so is its typical clientele. Their requirements and expectations differ greatly. Every business in this industry must meet each customer individually to be successful. All industries with such a wide target population use market segmentation since all customers want to be handled according to their preferences. Through this procedure, the entire client base is separated into groups with traits in common and, to a greater or lesser extent, similar requirements and demands. This could lead to the delivery of a far more specialist and customized service. Classification is done into ever-smaller groupings and is more accurate when utilizing machine learning techniques. Subgroups that had not been previously discovered could be included in this, along with the service is carefully tailored to each individual customer, the quality rises. This makes it possible for customers to have a better shopping experience, which makes them happier and raises the company's profitability. Machine learning is paving the path for technological advancement across all industries, having a significant impact on the travel and tourism market. These processes have a great deal of potential. There is little doubt that they can fundamentally alter how the travel industry operates and how commercialization takes place.

5. Comparison of Tourist Prediction Using Machine Learning Algorithms

Forecasting the amount of tourists coming to India, the study used certain well-known machine learning prediction algorithms in this part. The dataset from the website www.Kaggel.com was used for this investigation. A number of algorithms were used, including Naive Baysan, Logistic Regression, K - Nearest Neighbors, Naive Bayes, Support vector Regression, Random Forest, Log-Regression and Decision Tree Regression. The algorithms were contrasted based on their best predictions. A table showing the research's findings can be found below.

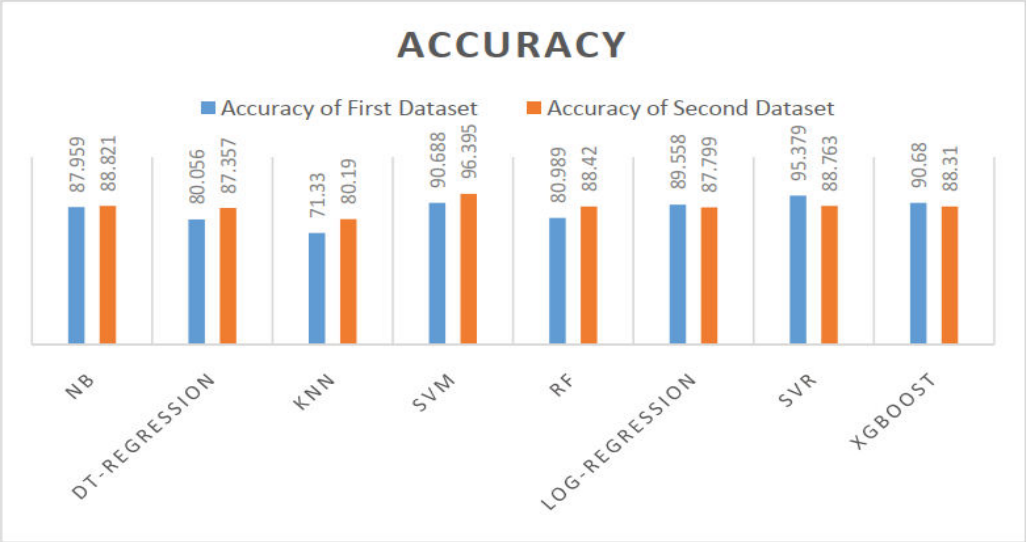
No	Algorithm	Accuracy	
		First Dataset	Second Dataset
1	**Naive-Bayes**	87.959	88.821
2	**Decision Tree**	80.056	87.357
4	**K - Nearest Neighbors**	71.330	80.190
5	**Support vector machine**	90.688	96.395
6	**Random Forest**	80.989	88.420
7	**Logistic Regression**	89.558	87.799

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105346 |

8	**Support vector Regression**	95.379	88.763
9	**X .Gboost**	90.68	88.31



V.CONCLUSION

In this work, the predictive abilities of seven algorithms is proposed—Decision Tree Regression, Support Vector Machine, K Nearest Neighbours, Naive Bayes, Support vector Regression, Random Forest, and Log-Regression were investigated. A data collection of 48 rows and 42 columns was used initially. The second dataset (q1), which comprised 12 rows and 42 columns, was integrated with the q1, q2, q3, and q4 datasets in order to estimate the number of tourists. India's annual visitor arrival data are included in the first dataset, and international visitor arrival data for the first quarter are included in the second dataset.

REFERENCES

1. Ghani, T., Jahan, N., Ridoy, S.H., Khan, A.T., Khan, S. and Khan, M.M., 2018, May. Amar Bangladesh-a Machine Learning Based Smart Tourist Guidance System. In *2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)* (pp. 1-5). IEEE.
2. Hapsari, I. and Surjandari, I., 2018, May. Visiting Time Prediction Using Machine Learning Regression Algorithm. In *2018 6th International Conference on Information and Communication Technology (ICoICT)* (pp. 495-500). IEEE.
3. Kbaier, M.E.B.H., Masri, H. and Krichen, S., 2017, October. A personalized hybrid tourism recommender system. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)* (pp. 244-250). IEEE.
4. Lin, D.J., Tsao, W.L., Yu, C.H., Liu, H.H. and Chang, Y.C., 2019, April. The Travel Time Prediction by Machine Learning Methods with Traffic Data in Chiayi City, Taiwan. In *2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT)* (pp. 257-260)
5. Nilashi, M., Bagherifard, K., Rahmani, M. and Rafe, V., 2017. A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & industrial engineering*, *109*, pp.357-368.
6. Son, K., Byun, Y. and Lee, S., 2018, October. Prediction of Visitors using Machine Learning. In *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (Vol. 3, pp. 138-139). IEEE.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN 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

www.ijircce.com

Scan to save the contact details