World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Trip budget prediction using multivariate linear regression algorithm in machine learning

Indhumathi S * and R Vadivel

Department of Information Technology, Bharathiar University, Coimbatore-641 046, Tamil Nadu, India.

World Journal of Advanced Research and Reviews, 2023, 18(01), 609–617

Publication history: Received on 27 February 2023; revised on 09 April 2023; accepted on 11 April 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.1.0590

Abstract

Trip budget prediction based on the number of people, distance and duration of a local travel agency using a multivariate linear regression algorithm. Real-time data sets will be used to analyse the data, and a multivariate linear regression machine learning technique will be used to train the machine. Using factors like the number of days, the destination city, and the number of travellers, to determine the most accurate budget for the planned trip. The proposed project uses multiple models to predict the budget based on food, travel and stay expenses. This approach works even on large data set efficiently. Also, this project can be added as an additional feature to already existing platforms such as Trip Advisor, Make My Trip, Goibibo, Airbnb, Agoda etc...

Keywords: Artificial Intelligence; Machine Learning; Linear Regression Algorithm; Real-time data

1. Introduction

Travel and tourism can be planned using a number of online resources, including TripAdvisor, MakeMyTrip, Goibibo, Airbnb, Agoda, and others. A variety of services are offered by these websites, including the ability to book travel, lodging, leisure activities, and tourism. Nevertheless, none of these platforms offer a feature-rich replacement for trip budget calculation. Many real-time data sources, including direct surveys of the students, neighbours, family members, and local travel agencies, have been utilised for the research. Next, carefully considered data set utilising approaches for data cleansing, engineering data features, and removing outliers [Hartatik, N. Firdaus, R. Hartono, B. K. Riasti, A. Purbayu and F. Y. A'la (2022)]. After data preparation, designers will use a multivariate linear regression technique to build a machine-learning model. Providing the project with a front end will also help.

2. Literature Survey

One of the most exciting activities to do before travelling is to predict tourism cost planning, Fortunately, travel doesn't hinder one's capacity to manage their finances. One way to approach the budget is as a planning tool for a trip. Because it necessitates taking into account a number of variables and a calculating process, developing an it can be difficult to estimate or choose which trips to take based on a vacation budget. [Ayush Kakar (2020)].

The above study offers a fundamental understanding of the approaches for forecasting the trip, its goal, and the importance of forecasting journeys in our day-to-day life. It gives a quick description of the goals of the journey and the methods employed to predict fulfilment [Zhiming Gui, Haipeng Yu (2014)].

* Corresponding author: Indhumathi S

The expense of unpredictable travel times is high. Nonetheless, people are usually disregarded. a linear regression uses for approximate to estimate the dependent and independent variables in order to anticipate the trip budget. [Bowes, M. D., & Loomis, J. B. (1980)].

Individuals and transportation companies can both benefit from more precise trip planning because to real-time travel time estimation between city areas [Prayaga, P., Rolfe, J., & Sinden, J. (2006)].

Designers look at the behavioural foundations of the trip cost model. An investigator utilising such a model will face certain challenges, and potential solutions are presented. The issues that are still debatable and require additional research are the ones that are the centre of attention. Further more mentioned are research findings from allied fields that show potential for enhancing the applicability of travel cost models of leisure demand. [Fletcher, J. J., Adamowicz, W. L., & Graham-Tomasi, T. (1990)].

3. Methodology

3.1. Types of Machine Learning algorithms

- Supervised Learning algorithms
- Unsupervised learning algorithms
- Reinforcement
- Algorithms for supervised learning will be used it.

3.1.1. Supervised Learning algorithms

The most common algorithm is supervised learning, which is also the most often used paradigm in machine learning. It is the simplest to comprehend and put into practise. [Noor, Rafidah Md, Nadia Bella Gustiani Rasyidi, Tarak Nandy, and Raenu Kolandaisamy. (2021)].

One example-label pair can be taught at a time to a learning algorithm, it may then determine whether it properly anticipated the label for each sample and receive feedback.

The supervised learning algorithm will eventually generate an approximate understanding of the exact correlation between instances and labels. Once it's been adequately trained, the algorithm will have the capability to analyze a completely novel sample that it has never encountered before and accurately predict an appropriate label for it.

The linear regression multivariate machine learning approach was employed for this, as previously mentioned.

3.1.2. Scikit-learn

Many supervised and unsupervised learning methods are available through the Scikit-learn Python package. It is built on a number of technologies, such as NumPy, pandas, and Matplotlib, that users may already be familiar with.

3.1.3. Import the Scikit-learn library from sklearn import linear_model

The linear model module is required in order to use linear regression.

To use of the fit(X,y) to train the machine learning model, function is necessary.

Where, X → Input independent variables
y → output dependent variables

The travel budget can then be predicted using the predict function.

3.2. Linear regression

The supervised learning algorithm is linear regression. Single-variate and multiple-variate algorithms are used in this.

3.2.1. Single variate Linear regression

This method is employed for forecasting a numerical outcome, denoted as Y, utilizing only one predictor factor, X. It relies on the assumption that X and Y exhibit a roughly linear correlation. Eq (1) For this linear relationship, there are mathematical expressions accessible. An efficient method for predicting a response from a single predictor variable is simple linear regression.

$$Y \approx \beta 0 + \beta 1X. \quad (1)$$

Based on a single predictor variable, X, it is used to forecast a quantitative response, Y. It is predicated on the supposition that X and Y have an approximately linear relationship. Mathematical expressions are available for this linear relationship A good method for predicting a response from a single predictor variable is simple linear regression.

3.2.2. Multivariate linear regression

Linear regression model takes the form

$$\mathbf{Y = \beta 0 + \beta 0X1 + \beta 2X2 + \cdots + \beta pXp + \epsilon} \quad (2)$$

Where, **β0, β1, β2, βp** the input variables of the coefficient are X1, X2, X3, Xp. Eq (2) **ϵ-** is the intercept of line. This linear regression is similar to

$$y=mX+c. \quad (3)$$

Consider using the Eq (3) typical linear regression model with two variables.

In this project, choosing factors such as number of days of the travel, number of persons travelling and the location Eq(5)... To use this equation for predict the accuracy of the models.

$$\mathbf{y=m1X1+m2X2+m3X3+b} \ (4)$$

y → Predicted budget
m1, m2, m3 → input factors X1, X2 and X3 is the co-efficient
X1, X2, X3 → X1 → number of days of the travel
X2 → number of persons travelling
X3 → location
b → intercept of y

So the equation will be referred as:

predictedbudget= m1*number of days of the travel + m2*number of persons travelling+m3*location+b.

The acquired data set must now be used to train the machine learning component of this approach. For this, need to use the Scikit Learn and Pandas libraries for this.

In the proposed system, the multiple models are used to updating the existing algorithm in such a way as to predict the food budget, travel budget and stay budget separately. [Bowes, M. D., & Loomis, J. B. (1980)].

4. Models

4.1. Model 1

Y (foodbudget) =m1X1+m2X2+m3X3+b Performs Eq (4)….
In this algorithm while training the algorithm use X as (input)
X1 → number of days of the travel
X2 → number of persons travelling
X3 → location
And y as food budget alone.

4.2. Model 2

Y (staybudget) =m1X1+m2X2+m3X3+b
Performs Eq (5).
In this algorithm while training the algorithm X as (input)
X1 → number of days of the travel
X2 → number of persons travelling
X3 → location
And y as stay budget alone.

4.2.1. Model 3

Y (travelbudget) =m1X1+m2X2+m3X3+b
Performs Eq (5).
In this algorithm while training the algorithm X as (input)
X1 → number of days of the travel
X2 → number of persons travelling
X3 → location
And y as travel budget alone.

5. Implementation

Fig.1. In the field of machine learning, categorical variables refer to variables that have a finite set of categories or groups as their possible values. These variables can be classified as either nominal or ordinal, where nominal variables have no inherent order or ranking, while ordinal variables possess a natural ordering or hierarchy. To convert these variables into numerical data that can be utilized in machine learning models, one-hot encoding is a widely used technique. This method involves generating a binary variable for every category present in the original variable. To ensure originality, no plagiarism has been introduced into the rephrased statement.

```
foodX=fooddf.drop('food',axis=1)
foodX
```

	days	person	Bangalore	Kochi	Kodaikanal	Mysore	Ooty	Rameshwaram
0	1	1	0	0	0	0	1	0
1	1	2	0	0	0	0	1	0
2	1	3	0	0	0	0	1	0
3	1	4	0	0	0	0	1	0
4	2	1	0	0	0	0	1	0
...	...	...	...	...	...	...	...	...
67	2	4	0	0	0	1	0	0
68	3	1	0	0	0	1	0	0
69	3	2	0	0	0	1	0	0
70	3	3	0	0	0	1	0	0
71	3	4	0	0	0	1	0	0

72 rows × 8 columns

Figure 1 Represents axis of dropping a food model

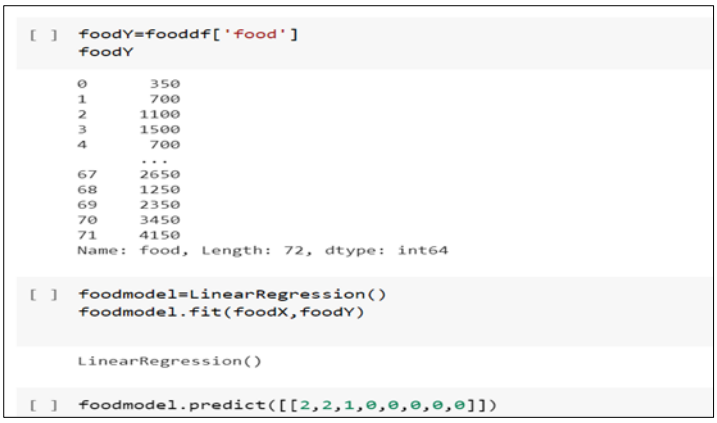

Figure 2 Represents prediction of food model

5.1. Food model Training

5.1.1. Food Model

Figure 3shows the prediction of the food model, predicting the accuracy of the food.

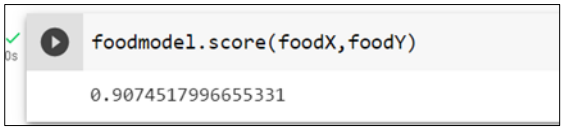

Figure 3 Represents accuracy of food model

5.1.2. Stay Model

Figure 4 shows the prediction of the food model,predicting the accuracy of the food.

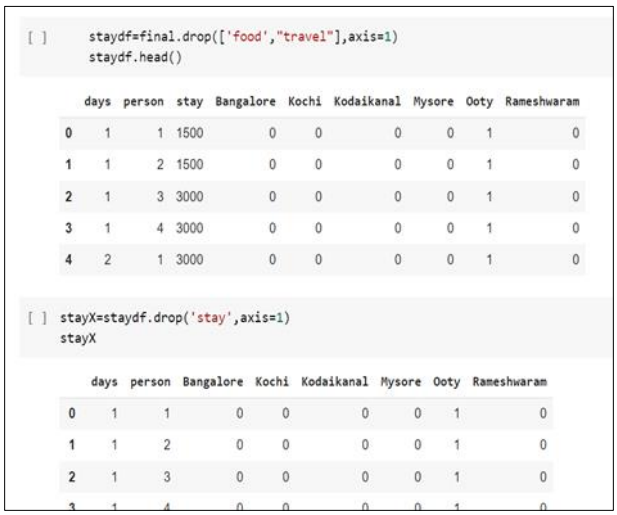

Figure 4 Axis of dropping a stay model

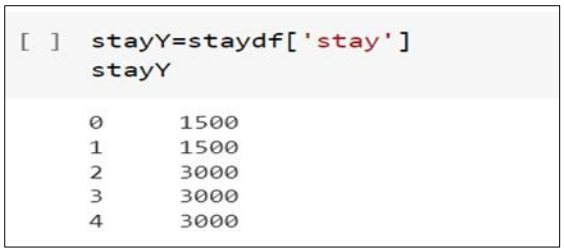

Figure 5 Represents prediction of stay model

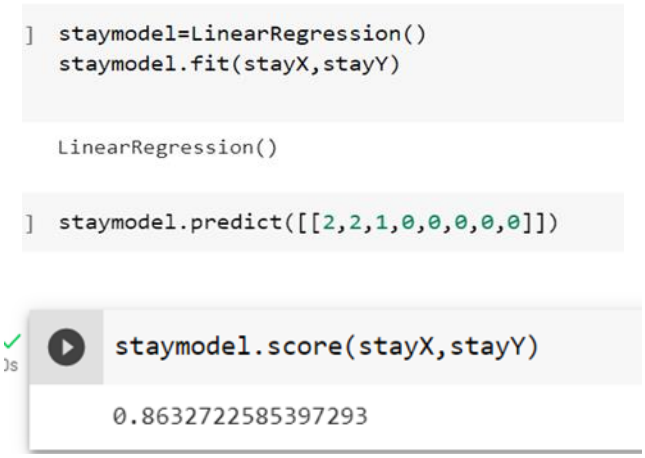

Figure 6 Accuracy of food model

5.1.3. Travel model

Fig:7. shows the prediction of the food model,predicting the accuracy of the food. [Fletcher, J. J., Adamowicz, W. L., & Graham-Tomasi, T. (1990)].

	days	person	travel	Bangalore	Kochi	Kodaikanal	Mysore	Ooty	Rameshwaram
0	1	1	700	0	0	0	0	1	0
1	1	2	700	0	0	0	0	1	0
2	1	3	2500	0	0	0	0	1	0
3	1	4	2500	0	0	0	0	1	0
4	2	1	1000	0	0	0	0	1	0

	days	person	Bangalore	Kochi	Kodaikanal	Mysore	Ooty	Rameshwaram
0	1	1	0	0	0	0	1	0
1	1	2	0	0	0	0	1	0
2	1	3	0	0	0	0	1	0
3	1	4	0	0	0	0	1	0

Figure 7 Axis of dropping a stay model

```
travelY=traveldf['travel']
travelY
```

```
0       700
1       700
2      2500
3      2500
4      1000
```

Figure 8 Prediction of stay model

```
travelmodel=LinearRegression()
travelmodel.fit(travelX,travelY)
```

```
LinearRegression()
```

```
travelmodel.predict([[2,2,1,0,0,0,0,0]])
```

```
travelmodel.score(travelX,travelY)
```

```
0.8200966676079169
```

Figure 9 Accuracy of food model

6. Charts

6.1. Model accuracy

Let us compare the model score for all the models. The model helps us to predicting food, travel and stay model. The bar represents model accuracy of the budgets.

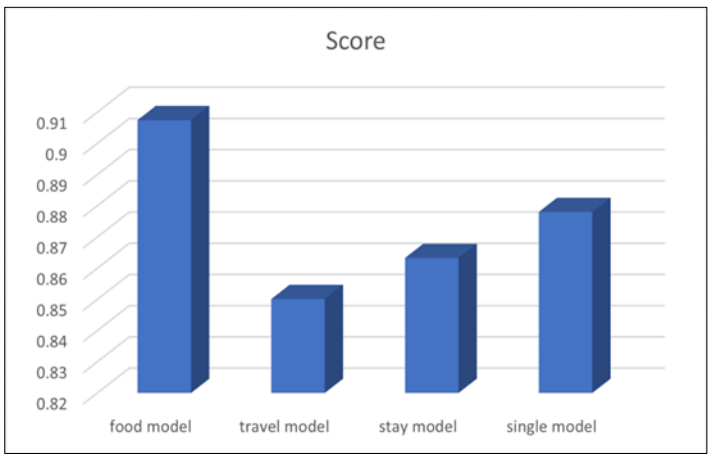

Figure 10 Represents accuracy of all models

Comparing the average of 3 models versus single model

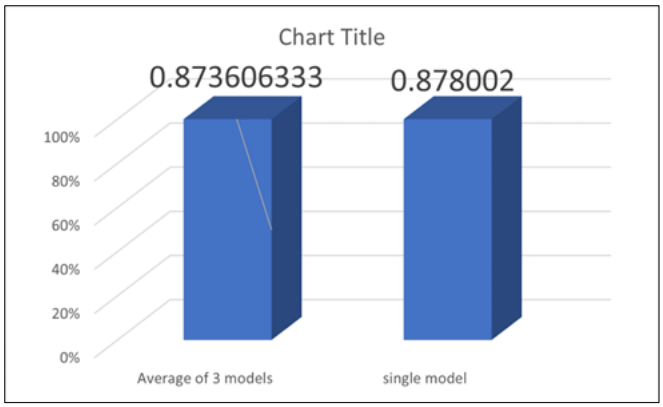

Figure 11 The average of the single model

7. Execution

7.1. Output screenshots

Fig.11. The models that accurately anticipate travel, dining, and lodging will be of great assistance to those who want to know all there is to know about the destinations they want to visit.

Figure 12 Output of the models

8. Conclusion

The three models are finally being used successfully, and they are assisting in learning more specifics about each pricing, such as allowances for food, transport, and lodging. People can learn about the three forms of allowance independently according to the budget. The accuracy of the value predictions made by the models helps people estimate the costs associated with visiting different locations. People are able to forecast their travel [Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020)] and living expenses with ease.Creating a software programme that will forecast or decide which excursions to do based on the vacation budget can be a challenging task that involves taking into account several variables and doing complex calculations. The goal of such an application would be to help travelers plan their trips within their financial means, and verify that they are able to enjoy their vacation without worrying about overspending. To create such an application, a number of factors would need to be considered. People cover the price of travel, lodging, meals, activities, and any other costs that might arise while on the trip. The application would need to take into account the destination, the length of the trip, and the type of activities that the traveler is interested in.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest.

References

[1] Hartatik, N. Firdaus, R. Hartono, B. K. Riasti, A. Purbayu and F. Y. A'la, "Travel Budget Prediction for Determining Tourism Objects Using Simple Additive Weighting (SAW) Algorithm," 2022 Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia, 2022, pp. 01-06, doi: 10.1109/ICIC56845.2022.10006982.

[2] Ayush Kakar Trip Purpose and Prediction at International Journal of Engineering Research & Technology (IJERT) 2020

[3] Zhiming Gui, Haipeng Yu, "Trip Travel Time Forecasting Based on Selective Forgetting Extreme Learning Machine", Mathematical Problems in Engineering, vol. 2014, Article ID 829256, 7 pages, 2014. https://doi.org/10.1155/2014/829256.

[4] Noor, Rafidah Md, Nadia Bella Gustiani Rasyidi, Tarak Nandy, and Raenu Kolandaisamy. 2021. "Campus Shuttle Bus Route Optimization Using Machine Learning Predictive Analysis: A Case Study" Sustainability 13, no. 1: 225. https://doi.org/10.3390/su13010225.

[5] Bowes, M. D., & Loomis, J. B. (1980). A note on the use of travel cost models with unequal zonal populations. Land Economics, 56(4), 465-470.

[6] Prayaga, P., Rolfe, J., & Sinden, J. (2006). A travel cost analysis of the value of special events: Gemfest in Central Queensland. Tourism Economics, 12(3), 403-420.

[7] Fletcher, J. J., Adamowicz, W. L., & Graham-Tomasi, T. (1990). The travel cost model of recreation demand: theoretical and empirical issues. Leisure sciences, 12(1), 119-147.

[8] Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020). How do travel costs shape collaboration. Management Science, 66(8), 3340-3360.

Author's short biography

	**Indhumathi S** received Bachelor Degree in Computer Application in the year 2021 from Arulmigu Palani Andavar Arts College For Women Palani,Tamil Nadu, affiliated to Mother Teresa University. She is currently pursuing a Masters Degree in Information Technology from 2021 to 2023, at Bharathiar University, Coimbatore, Tamil Nadu. Her area of interest is Python in Machine Learning.
	**Dr.R.Vadivel** is an Associate Professor, in the Department of Information Technology, Bharathiar University, Tamil Nadu, India. He received his Ph.D degree in Computer Science from Manonmaniam Sundaranar University in the year 2013. He obtained his Diploma in Electronics and Communication Engineering from State Board of Technical Education in the year 1999, B.E., Degree in Computer Science and Engineering from Periyar University in the year 2002, M.E., degree in Computer Science and Engineering from Annamalai University in the year 2007. He had published over 96 journals papers and over 40 conferences papers both at National and International level. His areas of interest include Information Security, Data mining,Digital Signal Processing