PREDICTION OF TARIFF RATE USING MULTIPLE LINEAR REGRESSION

Chevuru Nandini*,Dr.K.Reddy Madhavi

Original Article

Sree Vidyanikethan Engineering College, Tirupati, 517102, India

Corresponding Author's Email: nandinichevuru999@gmail.com and kreddymadhavi@gmail.com

Abstract

A tariff is a list of costs incurred during the movement of commodities across different distances. Seasonal and non-seasonal variables also affect tariffs. The goal of this study is to use multiple linear regression, a machine learning method, to estimate truck load tariff rates. We will attempt to forecast the truck tariff rates by using a few variables and the machine learning regression procedure previously discussed. By doing this, we can assist the industries in estimating the tariff rates so they can take the appropriate steps and manage and control the cost of transportation to run their businesses profitably.

Keywords: Machine learning, Multiple Linear regression, Tariff rate, Decision tree Regression, Random forest regression

Introduction

A tariff is a penalty placed by the government on products and services that are imported from foreign countries. Its goal is to increase the price of imports and lessen their attractiveness or competitiveness as compared to goods and services produced domestically. Tariffs are frequently implemented for limiting commerce with specific nations or to lower the importation of particular products and services. Factors used for tariff rate prediction are Seasonal Influences, Fuel Cost, Distance, Toll charge, Labor cost, Demand, Type of Highway, Weight of Load to be transported and Other Miscellaneous factors. Tariff rate is predicted by using Multiple linear regression algorithm and the tariff rate is displayed as the result using a web application. It is the most fundamental and widely utilized kind of predictive analysis. Through statistical modeling, the relationship between a dependent variable and a specified set of independent factors is modeled.

Literature Review

Many approaches for gathering tariff data have been developed recently in response to the increasing need for information on tariff rates. Instead of computing the tariff rate by hand, there are methods for predicting it that are more productive[7].Annual tariff data is collected from the past import and export records. Here we collected the data from kaggle. However, based on the number of factors we are considering the accuracy varies. So if we consider more number of factors the more the accuracy will be. However, we can also collect the data from other sources. The data from the above mentioned data sources have been gradually applied to estimate tariff rate. And here we considered the factors like Seasonal Influences, Fuel Cost, Distance, Toll Charge, Labor Cost, Demand, Type of Highway, Weight of Load to be transported and other miscellaneous factors. There are many techniques for predicting the tariff rate like lasso regression, simple linear regression, SVM etc. and out of all those techniques Multiple Linear Regression predicted the

result with more accuracy. There are also several research that concentrate on predicting short-term tariff rates using past tariff data[8]. Some writers, for example, used SVMs to predict the tariff rate for a given year based on the tariff rate from previous years[9].Additionally, several writers addressed the distinctions and overlaps between neural networks and statistical techniques in the domain of transportation, specifically in relation to forecasting and tariff rate analysis[10]. When utilizing these strategies in comparison to standard models, improved results were observed. In this study, the assessment of the tariff rate took into account all relevant parameters, including demand, kind of roadway, fuel cost, distance traveled, tolls, labor costs, seasonal effects, and weight of the load to be transported, among other unspecified factors.This work use for the Linear regression, Lasso regression, Ridge regression, and Elastic net regression to filter the raw data set and analyze the regression model in order to forecast the cost model[1].This study uses feature extraction, data acquisition, model training, and model deployment to forecast the tariff rate for truck loads[2]. The Benchmarking is less than truckload market rates are estimated in this article. It uses regression techniques and formulates multiple regression to estimate LTL price[3].The goal of this research is to determine the best household load control in a real-time electricity pricing scenario by predicting prices. The price prediction problem and the automation systems are fully utilized[4]. POU pricing based on the ideal load profiles for each aggregator. It takes into account the erratic nature of the arrival and departure timings as well as the energy consumption of PEVs linked to each aggregator transformer, hence reducing the transformers' aging rate[5].The SAE model is shown. One well-known deep learning model is the stack of auto encoders, or SAE model. It builds a deep network latent traffic flow feature representation, like the nonlinear one, using auto encoders as building blocks[6].

Proposed Approach

The main purpose of our paper is to Predicting of the Tariff rate for truck carrying goods. The system is designed using one of the machine learning techniques which uses historical data for developing a Multiple Linear Regression based model and is capable of predicting the tariff rate for new truck carrying the goods. The system must be able to fetch the values like fuel rate, climatic conditions etc. dynamically and they must be used for predicting the values.

This paper is aimed to automate the process of predicting the tariff rates on trucks while transporting the goods from one place to another place. The predicting system includes the following: 1. Ability to predict tariff rate for goods based on historical data by using multiple linear regression machine learning algorithm.

Multiple Linear Regression:

The Linear relationship between a single continuous dependent variable and several independent variables is modeled using the significant regression procedure known as multiple linear regression. Given that several predictor variables are required to predict the response variable, multiple linear regression is an extension of simple linear regression. The predictor or independent variable for multiple linear regression (MLR) can be continuous or categorical, but the dependent or target variable (Y) must be continuous or real. Every feature variable needs to simulate the dependent variable's linear connection. Using a multidimensional space of data points, MLR attempts to fit a regression line. The MLR formula

Multiple predictor variables, x_1, x_2, x_3,..., x_n, are combined linearly to form the target variable (Y) in multiple linear regression (MLR). Considering that multiple linear regression is an improvement over simple linear regression, the same rules apply. Equation 1 is

$$\mathbf{Y= b_0+b_1x_1+b_2x_2+\ldots..b_nx_n} \tag{1}$$

Where as,

Y= Response variable/output,

b_0, b_1, b_2,….,bn= Coefficients of the model.

x_1, x_2, x_3,...,x_n forms various Independent/feature variable.

IJIEE

Multiple Linear Regression Pseudocode

Input:

- X - Independent variables matrix (features)
- y - Dependent variable vector

Output:

- Coefficients b0, b1, ..., bn

Function to perform multiple linear regression

function multiple_linear_regression(X, y):

- **Step 1: Initialize variables**

n_samples, n_features = shape(X)

ones_vector = ones(n_samples)

X_with_intercept = concatenate(ones_vector, X, axis=1) # Add a column of 1s for the intercept

- **Step 2: Calculate coefficients using least squares method**

coefficients = inverse(transpose(X_with_intercept) * X_with_intercept) * transpose(X_with_intercept) * y

return coefficients

- **Example usage:**
    - X_train - Training data features
    - y_train - Training data labels

coefficients = multiple_linear_regression(X_train, y_train)

print("Coefficients:", coefficients)

Procedure:

Step 1: Historical Data of Tariff rate is collected.

Step 2: Machine learning model is built using Multiple linear Regression Model by splitting the dataset into 2 parts. one is for training the model and other is for testing the model.

Step 3: Save the model to disk by using pickle library and corresponding script.pkl file is generated

Step 4: Load the script.pkl file in app.py and define routing parameters for rendering HTML file

Step 5: New values are given as the input for predicting the new Tariff rate.

Step 6: Display the Predicted tariff rate on screen using web application.

Results and Discussion

Machine Learning model is trained with some data set before providing to the user. This data set consists of all the factors using which tariff rate is predicted. After the model is trained with the dataset testing need to be done with the testing data set. Here we used Multiple Linear Regression Algorithm to predict the tariff rate. The coefficients ($b0,b1,…,bn$) derived from the multiple linear regression model reveal the expected change in the tariff rate for a one-unit change in the corresponding independent variable, assuming other variables remain constant. The intercept ($b0$) represents the estimated tariff rate when all independent variables are zero. The model's performance is quantified using

the Mean Squared Error (MSE) as specified in table 1, which measures the average square difference between the predicted and actual tariff rates. The calculated MSE is X (insert the actual MSE value here). A lower MSE signifies a closer fit between predicted and observed values. P-Values are associated with each coefficient indicate their statistical significance. A p-value less than 0.05 is suggests to the corresponding independent variable significantly influences the tariff rate. Figure 1 specifies the actual versus predicted values for the proposed method. The plot for multilinear regression is shown in figure 2. Figure3 and Figure4 represents the decision tree and random forest regression models.
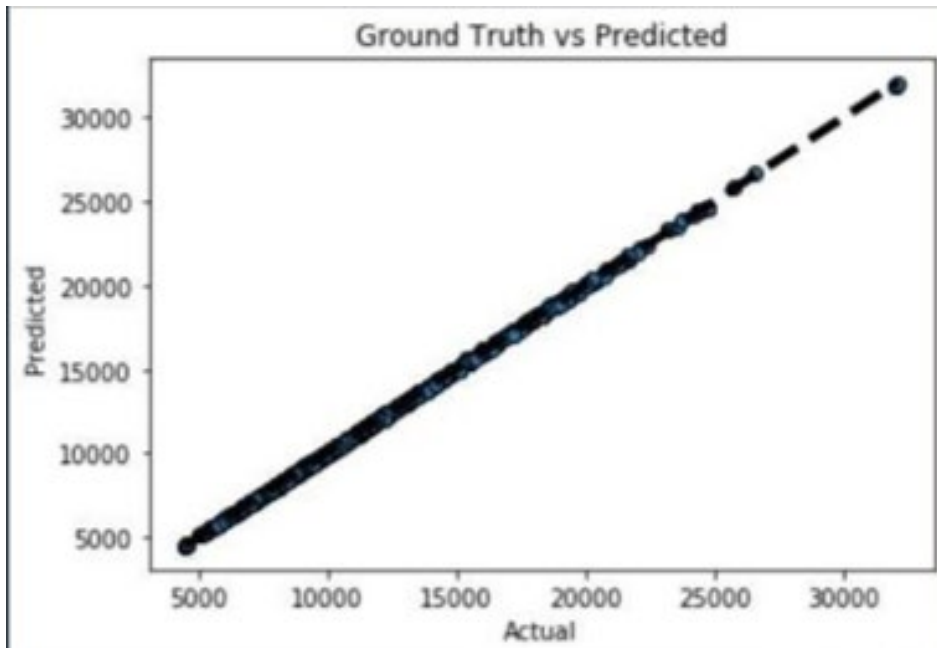
Figure 1 Specifies the Ground Truth VS Predicted Values
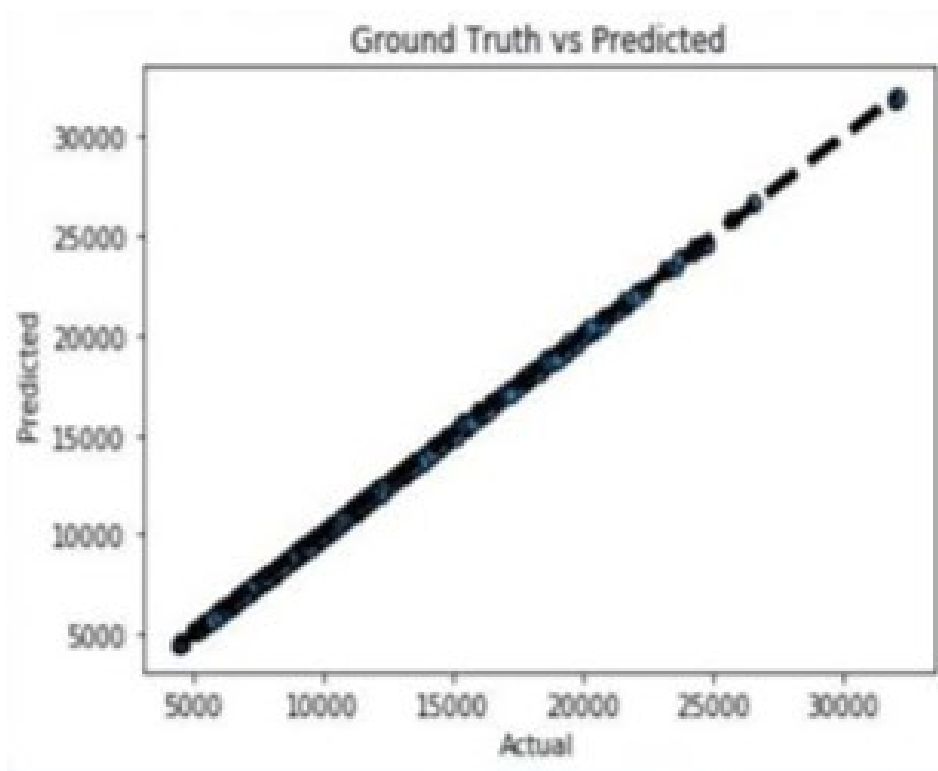

Figure 2 Plot for Multi Linear Regression
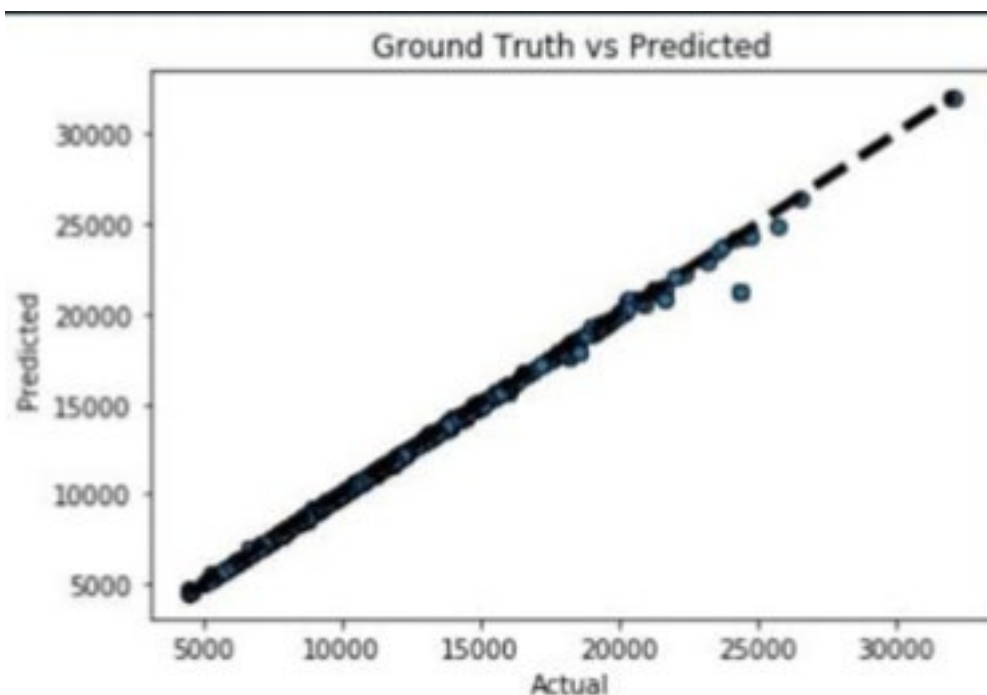
Figure 3 Plot for Decision Tree Regression

Figure 4 Plot for Random Forest Regression

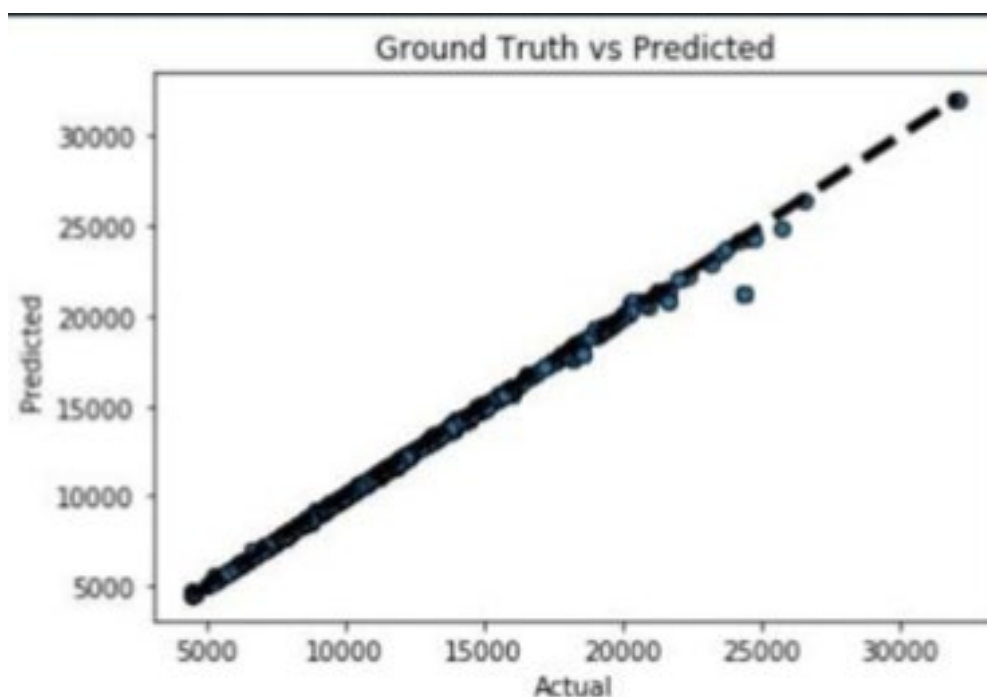*Table 1 Comparison of MSE and Test Variance Score*

APPROACH	MEAN SQUARE ERROR	TEST VARIANCE SCORE
Multiple Linear Regression	2834.4812565441525	0.9998509684905079
Decision Tree Regression	48638.30465949821	0.9974426925752978
Random Forest Regression	43657.0108850019	0.997704599309986

Conclusion & Future Work

We have managed out a way to prepare a model that offers users a best approach with take a glance at future worth predictions. By exploitation the info that we have a tendency to gathered from the web we showed that it's possible to predict the tariffs rates that helps in many ways that anyone will predict the tariff rates through all tolls and seasonal variations. As we have predicted the tariff rate tax using Multiple Linear Regression similarly we could be able to predict the other government taxes. This model can be developed into a cross platform for all the other government taxes by applying more machine learning algorithms and by increasing the accuracy of the model. The Graphical user interface can also be developed into a multiple page interface. We hope that the presented model can be successfully used in the practice as very accurate, efficient and fast forecasting model for prediction of tariff rate. By considering some more additional factors we can improve accuracy of the result predicted. By using Support Vector Machine Regressor Model, we can achieve more accuracy. Here we used MLR for simplicity.

References

1. Sree Ram Murthy, K., Kranthi Kumar, K., Sai krishna, D., Sai Prasad, D., Manideep Reddy Dara, Divya, G.: Prediction of Tariff Rate. International Journal of Advanced Science and Technology, vol. 07, no. 05, May (2020).
2. Arulkumar, V., Agalya, S., Gokul Chakkaravarthi, S., Gokula Priya, E.: Tariff Rate Prediction for Truckload Based on Predictive Analysis using Machine Learning. International Journal of Advanced Science and Technology, vol. 29, no. 9s, p. 5477 5482, (2020).
3. Evren Ozkaya., Pınar Keskinocak., Roshan Joseph, V., Ryan Weight.: Estimating and benchmarking Less-than-Truckload market rates, Transportation Research, p 667-682, (2019).
4. Amir Hamed., Mohsenian Rad., Alberto Leon Garcia.: Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environment, IEEE TRANSACTIONS ON SMART GRID, vol. 1, no. 2 (2010).
5. Sima Davarzani., Ioana Pisica., Gareth, A.: Assessment of Residential Electricity Tariff Switching Based on Customer Response Elasticity, Institute of Energy Futures, Smart Power Networks Brunel University London, (2016)
6. Kiguchi, Y., Y. Heo, M. Weeks, and R. Choudhary. "Predicting intra-day load profiles under time-of-use tariffs using smart meter data." Energy 173 (2019): 959-970.
7. Shyamala Devi, M., Ganta Umamaheswar Reddy, Balla Teja Swaroop, and Rachaputi Kiran Kumar. "Regressors with Anova-Reduced Features for Tariff Rate Prediction Using Machine Learning." In Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021, pp. 317-326. Singapore: Springer Nature Singapore, 2022.
8. Milad, Mohamed AH, Rose Irnawaty Ibrahim, and Samiappan Marappan. "Regression Model Building and Forecasting on Imports in Malaysia." Australian Journal of Basic and Applied Sciences (2015).
9. Wu, Huaiguang, Mingxing Zhang, and Baohua Jin. "Application of an Ensemble Learning Method in the Risk Prediction of Electricity Tariff Recovery." In Journal of Physics: Conference Series, vol. 1549, no. 5, p. 052126. IOP Publishing, 2020.
10. Chellamani, Ganesh Kumar, M. Firdouse Ali Khan, and Premanand Venkatesh Chandramani. "Supervised electricity tariff prediction using random forest validated through user comfort and constraint for a home energy management scheme." Journal of Intelligent & Fuzzy Systems 40, no. 1 (2021): 745-757.