Agriculture Consignment Transporter with ML Techniques

Sivandan S,K.S.Giriprasath
M.Sc. Data Science and Business Analysis, Department of Computer Science
Rathinam College of Arts and Science,Coimbatore

Abstrac—Precise estimation of costs for agricultural consignment shipping is essential for logistics providers as well as farmers. In order to preserve their competitiveness and financial stability, logistics providers must maintain accurate pricing, while farmers must assess transport costs in order to determine profit margins. Conventional approaches frequently depend on set prices or straightforward computations, which ignore the dynamic character of the agricultural transportation industry. In order to anticipate agricultural consignment transport prices with greater flexibility and accuracy, this paper presents a machine learning (ML) model. By capturing the intricate interactions between the several elements driving transit costs, the model seeks to help both parties make more educated price decisions. This work offers a viable method for utilizing ML algorithms to forecast agricultural consignment shipping pricing. There is a chance that the suggested model will enhance pricing, precision, adaptability, and transparency, which are advantageous to agricultural supply chain logistics suppliers as well as farmers.

***Keywords*: Machine Learning, Auto ML, Auto SK Learn, Agriculture Consignment Transporter (Act) Fare Detection.**

1. INTRODUCTION

The foundation of economies around the world is the agricultural sector, which supplies food and raw materials to a wide range of sectors. In this industry, the smooth movement of goods from fields to markets is largely dependent on the transportation of agricultural consignments. The unpredictability of transportation prices, which are impacted by factors like distance, weight of goods, and external economic variables like fuel prices, is one of the major issues faced by both farmers and consignment transporters in the agriculture consignment sector.

Transportation costs constitute a substantial portion of the overall logistics expenses in the agricultural supply chain. Accurate prediction of these costs is essential for consignment transporters to optimize their pricing strategies, plan routes efficiently, and enhance overall operational efficiency. For farmers, understanding transportation costs are crucial for budgeting and selecting cost-effective transportation options. Traditional methods of pricing estimation often lack precision and fail to account for the dynamic nature of the factors influencing transportation costs.

The main goal of this project is to create a predictive model that accurately estimates transportation costs for agricultural consignments by using machine learning techniques. The model attempts to provide consignment transporters, farmers, and other stakeholders with a dependable tool for more accurate and efficient transportation cost prediction by utilizing historical data and taking into account numerous pertinent criteria. The methodology involves collecting and preprocessing a comprehensive dataset encompassing historical transportation costs, distance travelled, types of goods transported, and other pertinent variables. Data preprocessing includes handling missing values, outliers, and ensuring data quality. Feature engineering techniques are employed to extract meaningful information from the dataset, enhancing the model's ability to capture the underlying patterns in the data. A number of machine learning techniques are explored and emphasized, including decision trees, random forests, support vector machines, and linear regression. To maximize process efficiency, hyperparameter adjustments are made after the model has been trained on a portion of the data. The model's ability to generalize to new data is evaluated using cross-validation procedures.

An essential aspect of this research is the emphasis on model interpretability. The goal is not only to predict transportation prices accurately but also to provide insights into the factors driving these predictions. A transparent and interpretable model enables stakeholders to understand the model's decision-making process, fostering trust and facilitating informed decision-making for farmers.

This paper based on using machine learning approaches to create an intelligent farm consignment transporter system. The system will prioritize dynamic pricing, resource allocation, and route optimization.

through keeping an eye on conditions, making sure perishables are handled properly, and guaranteeing prompt delivery. granting real-time visibility into the transportation process to all farmers and drivers. Lowering carbon emissions can be achieved in part by streamlining routes and cutting back on wasted mileage.

To evaluate the created model's performance on unobserved data, it is thoroughly validated using a different test dataset. Further validation of the model's adaptability and efficacy in real-world scenarios is done using real-world data from farm consignment transporters. In order for the model to be easily incorporated into farmer and consignment carrier’s daily operations and to function consistently under a variety of scenarios, it must first undergo validation.

II. LITERARY REVIEW

Recent study on the topic of vehicle fare prediction has aimed at developing data-driven approaches for forecasting future vehicle prices and their trends.

1. Ratnakanth G [1] utilised Deep Neural Network that functions same as the human brain. The data is preprocessed, and the Min-Max normalisation approach was used to change the values that are already present in the dataset in order to obtain excellent performance. Randomised Search CV algorithm is used for hyperparameter tuning of the Deep Learning algorithm. Finally, the dataset was visualised using univariate analysis, bivariate analysis, and correlative analysis for all of the features in the dataset.

2. "A Deep Learning Framework for Cab Fare Prediction" by Xia et al [2]. (2019) - This study proposes a deep learning framework for cab fare prediction, which uses historical fare data and traffic data to make predictions.

3. "Forecasting Cab Fare Prices Using Time Series Analysis and Artificial Neural Networks" by Singh et al [3]. (2017) - This study uses time series analysis and artificial neural networks to make predictions of cab fare prices and compares the performance of different neural network models.

4. "Cab Fare Prediction Using Ensemble Machine Learning Techniques" by Singh et al [4]. (2019) - This study uses ensemble machine learning techniques such as bagging, boosting, and stacking to make predictions of cab fare prices and compares the performance of the different ensemble methods.

5. S. Naveen Prasath et al [5] researched and found out the factors that impact the flight fare fluctuations. The paper systematically demonstrates the K-Nearest Neighbours technique to estimate the prices at a particular instance using Machine Learning techniques. After doing a comparison of the highest and lowest levels of airfare for specific days, weekends, and times of the day, such as morning, evening, and night, regression analysis was carried out to predict the flight prices.

6. M. Basyir, M. Nasir, and W. Mellyssa,[6] “Determination of Nearest Emergency Service Office using Haversine Formula Based on Android Platform’’.

7. R. Raja Subramanian et al [7] collected data from MakeMyTrip, Data World and Kaggle to build Machine Learning models. The paper uses KNN Regression, Linear Regression, Lasso Regression, Ridge Regression, and Random Forest Regression. The models have been implemented using the sci-kit learn python library. The research found out the Random Forest Regressor algorithm works the best with high accuracy.

III. METHODOLOGY

A. Importing Libraries and Dataset

The first step is importing different libraries that will help in feature extraction, data analysis and model building. Pandas [22] and NumPy [21] are installed in the Jupyter Notebook [30]. Matplotlib [24] and Seaborn [23] libraries are installed for visualisations. Next, two datasets consisting of training data and test data are installed in the Jupyter Notebook.

The testing data and the training data have been merged to produce a single dataset in order to make the process of Feature
Engineering more straightforward. The chosen dataset contains both training and testing data, contain different features such as the total distance, weight, and the Vehicle type.

B. Data Analysis

For this purpose, only the training dataset is used to perform data analysis. First, the correlation between Independent and dependent attributes is found out by using heatmap which is present in the seaborn library [23]. Figure 1 displays the results obtained in the form of data visualisation using heatmap.

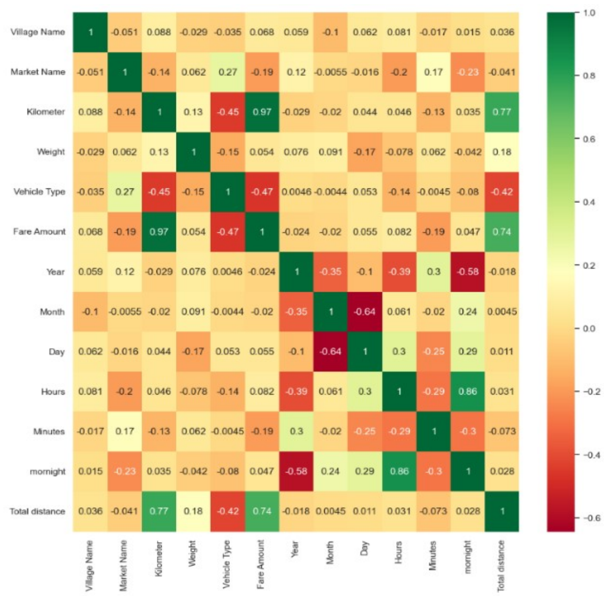

Figure 1. correlation between independent and dependent variable

C. Feature Engineering

The data must now be transformed into a format that the model can comprehend. For the model's interpretation, several forms of alphabetical or continuous data should then be transformed into numeric data. The Pickup Date column in the dataset contains data in the format of YYYY/MM/DD. For simplicity, the Pickup Date column is converted into 3 columns - Date, Month and Year. The original column is dropped. Then, the data type of the previous 3 columns is converted to integer type using the DataFrame.astype() method [28] function.

The Pickup time is not of consistent format throughout the dataset and may include both time and date or solely time. For this reason, the column is structured to display only the time in the format HH: MM. The Pickup time is then divided into 2 columns - Hour and Minute. Additionally, the data in the column that contains information about the total number of weights has been reformatted such that it will appear as a numeric value with an integer datatype. Similarly, all the other columns will be converted to numeric values.

The dataset includes a column labelled "weight," which displays the path that the vehicle takes. If it is a vehicle without any goods in between, then the weight will just provide the village name and the market name. Figure 3 displays the weight column in the dataset.

Vehicle Type	Village Name	Market Name	Weight
Tata 407	Maniyampattu, Tamil Nadu	Kanai	2500 kg
Tata ape	Aariyur, Tamil Nadu	Kanai	500 kg
Tata ape	Agaram Chithamoor	Kanai	438 kg
Tata 407	Ariyalur Thirukkai	Kanai	1500 kg
Tata 407	Kalpattu	Kanai	2500 kg

Figure 2. Dataset

In the test data, the price column is empty and has null values. This research paper considers both the test data and train data together. For time being, the price column in test data is filled with the mean of the price column in the train data. To use the label encoder [29], it needs to be imported from sklearn.preprocessing. Label encoding is used on all the columns of the dataset. The method fit_transform is used which is the combination of fit method and transform method, it is equivalent to fit().transform(). If we use fit and transform separately when we need both, then it will decrease the efficiency of the model so we use fit_transform() which will do both the work. The data is arranged categorically in a systematic manner. All the columns have been converted into labels.

D. Harversine Formula

In this project the data contains only the pick-up and drop points in longitude and latitude. The fare amount will mainly depend on the distance covered between these two points. which is a numeric value and explains the distance covered between the pick-up and drop of points. After researching I found a formula called haversine formula [6], that determines the distance between two points on a sphere based on their given longitudes and latitudes. This formula calculates the shortest distance between two points in a sphere.

Used in Python :

```
def haversine(df):
lat1=np.radians(df["Pickup Latitude"])
lat2= np.radians(df["Dropoff Latitude"])

### Based on the formula x1=drop_lat,x2=dropoff_long
dlat=np.radians(df["DropoffLatitude"]-df["Pickup Latitude"])
dlong = np.radians(df["Dropoff Longitude"]-df["Pickup Longitude"])
a = np.sin(dlat/2)**2 + np.cos(lat1) * np.cos(lat2) * np.sin(dlong/2)**2
c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1-a))
r = 6371 # Radius of earth in kilometers if we in miles let's use 3956
return c * r
```

After executing the haversine function in our project, I got new variable Total distance and some instances of data are mentioned below.

	Village Name	Market Name	Kilometer	Weight	Vehicle Type	Year	Month	Day	Hours	Minutes	mornight	Total distance
0	91	1	55	41	2	2015	1	27	9	8	0	32.012835
1	0	1	7	86	4	2015	1	27	9	8	0	2.525760
2	3	1	5	80	4	2011	10	8	7	53	0	3.126972
3	10	1	11	14	2	2012	12	1	17	12	1	7.692249
4	40	1	8	41	2	2012	12	1	17	12	1	5.758027

Figure 3. New variable total distance

E. Model Building using ML

The model is built using the Random Forest Regression [26] Machine Learning algorithm. A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. It is of two types - Random Forest Classifier and Random Forest Regressor. This

paper uses Random Forest Regressor.
The random forest model is a type of additive model that makes predictions by combining decisions from a sequence of base models. and the equation is

$$g(x)=f0(x)+f1(x)+f2(x)+...$$

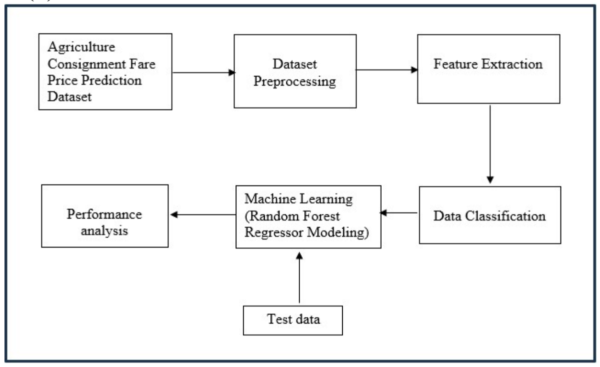

Figure 4. System Architecture Model

This paper discusses the use of Random Forest because this algorithm takes less time to train. It also predicts output with high accuracy and even for the large dataset, it runs efficiently. Random Forest Regressor can maintain accuracy even when a large proportion of the data is missing. The features selected for training are vehicle type, pickup location, drop location, Date, Month, Year, weight, and so on. Distribution plot of the difference between the actual and predicted values is created. From Figure 5, we observe that zero is in abundance and the plot follows normal distribution. This suggests that the difference between predicted value and actual value is zero. This ensures good accuracy of the project.

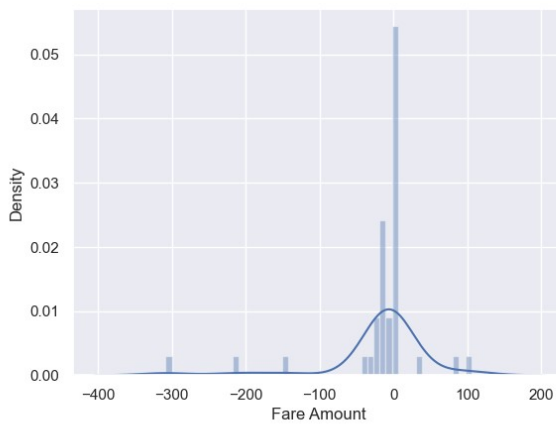

Figure 5. Distribution plot

F. Model Building and Prediction using Auto SK Learn

Automated Machine Learning will automate all the machine learning model building and hypertuning parts. Auto SK Learn is used for this purpose in this project. Auto SK Learn only automates the model building part. The preprocessing part needs to be done manually. If raw data is passed in the model without any preprocessing, it will cause the model to fail. This is because there will be different types of data available - unstructured, numerical, object, alphabetical, etc. After performing Auto ML, the top perfectly fitted models are found out. In this case, gradient boosting and random forest were the top models. Figure 6 displays the scatter plot obtained of the predicted and testing data. The cluster in the scatter plot means that the model has high accuracy and the prediction is performed successfully.

IV. RESULTS AND INFERENCES

The Random Forest Regressor algorithm has a mean absolute error of 28.534, mean squared error of 4469.7378425 and root mean square error of 66.8560. Coefficient of determination, also called as R^2 score is used to evaluate the performance of a regression model. It is the amount of the variation in the output dependent attribute which is predictable from the input independent variable(s). A higher value of R^2 is desirable as it indicates better results. The model built gives R^2 value of 0.97.

```
print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
print('MSE:', metrics.mean_squared_error(y_test, y_pred))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))

MAE: 28.222500000000004
MSE: 4158.946555
RMSE: 64.48989498363291

# RMSE/(max(DV)-min(DV))

2090.5509/(max(y)-min(y))

0.38044602365787084

metrics.r2_score(y_test, y_pred)

0.9729586466626785
```

Figure 7. Output

One paper utilises Deep Learning techniques such as Deep neural network [1]. The paper suggests that the accuracy obtained by using Deep learning is better than the accuracy obtained using Machine learning models. One of the most common limitations of this project is obtaining information because data is acquired from websites that book Agriculture consignment transporter.

The paper titled, "Airline Fare Prediction Using Machine Learning Algorithms" predicted a root mean square error of 33.36 when Random Forest Regressor algorithm is used [7]. The model used in the paper is hypertuned so that the error is reduced.

V. CONCLUSION AND FUTURE WORK

The research paper depicts how using Automated Machine Learning saves the time of model building but highlights that the data preprocessing part be done manually and that it cannot be automated. The prediction of the consignment rate was carried out successfully using one of the most widely used algorithms - Random Forest Regressor. The accuracy achieved is very high which is seen from the distribution plot and scatter plot obtained from the training data and testing data. The data visualisation techniques have been applied to illustrate the ideology behind the attributes of the dataset. To acquire more reliable findings, more accurate data with greater features might be employed. In future I decided to convert web application to mobile application for more user convenience.

REFERENCES

[1] G. Ratnakanth, "Prediction of Flight Fare using Deep Learning Techniques," 2022 International Conference on Computing, Communication and Power Technology (IC3P), 2022, pp. 308-313, doi: 10.1109/IC3P52835.2022.00071.

[2] “A Deep Learning Framework for Cab Fare Prediction” by Xia et al. (2019) - This study proposes a deep learning framework for cab fare prediction, which uses historical fare data and traffic data to make predictions.

[3] “Forecasting Cab Fare Prices Using Time Series Analysis and Artificial Neural Networks” by Singh et al. (2017) - This study uses time series analysis and artificial neural networks to make predictions of cab fare prices and compares the performance of different neural network models.

[4] Cab Fare Prediction Using Ensemble Machine Learning Techniques" by Singh et al. (2019) - This study uses ensemble machine learning techniques such as bagging, boosting, and stacking to make predictions of cab fare prices and compares the performance of the different ensemble methods.

[5] S. N. Prasath, S. Kumar M and S. Eliyas, "A Prediction of Flight Fare Using K-Nearest Neighbors," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 1347-1351, doi: 10.1109/ICACITE53722.2022.9823876.

[6] M. Basyir, M. Nasir, and W. Mellyssa, “Determination of Nearest Emergency Service Office using Haversine Formula Based on Android Platform,” Emit. Int. J. Eng. Technol., vol. 5, no. 2, pp. 270–278, 2017.. .

[7] R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy and R. R. Sudharsan, "Airline Fare Prediction Using Machine Learning Algorithms," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 877-884, doi: 10.1109/ICSSIT53264.2022.9716563.

[8] S. N. Anwar, I. Nugroho, and E. Supriyanto, “Model Rute dan Peta Interaktif Posyandu di Kota Semarang Menggunakan Geolocation dan Haversine Berbasis Mobile Android,” in PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU&CALL FOR PAPERS UNISBANK (SENDI_U), 2015.

[9] R. H. D. Putra, H. Sujiani, and N. Safriadi, "Penerapan Metode Haversine Formula Pada Sistem Informasi Geografis Pengukuran Luas Tanah," J. Sist. dan Teknol. Inf., vol. 1, no. 1, 2015.
[10] C. Chariton and Min-Hyung Choi, "Enhancing usability of flight and fare search functions for airline and travel Web sites," International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004., 2004, pp. 320-325 Vol.1, doi: 10.1109/ITCC.2004.1286473.
[11] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.
[12] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.
[13] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State
Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.
[14] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
[15] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.
[16] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.
[17] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756
[18] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749
[19] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique
with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
[20] M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
[21] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area
of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019
[22] Zhao, Z., You, J., Gan, G., Li, X. & Ding, J. 2022, "Civil airline fare prediction with a multi-attribute dual-stage attention mechanism", Applied Intelligence, vol. 52, no. 5, pp. 5047-5062.
[23] Malighetti, P., Paleari, S. & Redondi, R. 2009, "Pricing strategies of lowcost airlines: The Ryanair case study", Journal of Air Transport Management, vol. 15, no. 4, pp. 195-203.
[24] Tziridis, K., Kalampokas, T., Papakostas, G.A. & Diamantaras, K.I. 2017, "Airfare prices prediction using machine learning techniques", 25th European Signal Processing Conference, EUSIPCO 2017, pp. 1036.
[25] Groves, W. & Gini, M. 2013, "An agent for optimizing airline ticket purchasing", 12th International Conference on Autonomous Agents and Multiagent Systems 2013, AAMAS 2013, pp. 1341.
[26] Makridakis, S., Spiliotis, E. & Assimakopoulos, V. 2018, "Statistical and Machine Learning forecasting methods: Concerns and ways forward", PLoS ONE, vol. 13, no. 3.
[27] Lu, M., Zhang, Y. & Lu, C. 2021, "Approach for Dynamic Flight Pricing Based on Strategy Learning", Dianzi Yu Xinxi Xuebao/Journal of Electronics and Information Technology, vol. 43, no. 4, pp. 1022-1028.
[28] Joshi, N., Singh, G., Kumar, S., Jain, R. & Nagrath, P. 2020, Airline Prices Analysis and Prediction Using Decision Tree Regressor.
[29] Boruah, A., Baruah, K., Das, B., Das, M.J., Gohain, N.B. (2019). A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter. In: Panigrahi, C., Pujari, A., Misra, S., Pati, B., Li, KC. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol 714. Springer, Singapore.
[30] B. Purmadipta, H. Anra, and M. A. Irwansyah, "Sistem Informasi Geografis Perumahan dan Fasilitas Sosial Terdekat dengan Metode Haversine Formula," J. Sist. dan Teknol. Inf., vol. 1, no. 1, 2016.
[31] J. Carroll and S. Hughes, "Using a video camera to measure the radius of the Earth Joshua," Phys. Educ., vol. 48, no. 6, pp. 731–735, 2013.