International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n2.05
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Study on Illegal Wildlife Trade Based on Hierarchical Analysis and Random Forest Modeling

Ruiqi Zhou *

Delaware Institute of Data Science, Southwestern University of Finance and Economics, Chengdu, China

*Corresponding Author: 13379941615@sohu.com

ABSTRACT

With the goal of solving the problem of illegal wildlife trade, this study proposes a data-driven five-year project by combining hierarchical analytical modeling and random forest modeling. First, 16 potential factors were analyzed through a hierarchical analytical model to identify key customers. Then, a linear regression model was used to develop a five-year program suitable for the client. Subsequently, a random forest model was used to verify that the program met the client's needs, and model training was used to predict the impact and probability of success of the program. Finally, in conjunction with the AHP model, the program was confirmed to significantly reduce illegal trade. This study proposes an integrated approach that provides an effective solution for wildlife conservation, and hopefully will lead to a win-win situation for the protection of ecological balance and human interests.

KEYWORDS

AHP; Linear Regression Modeling; Random Forest

1. INTRODUCTION

Illegal wildlife trade, the fourth largest illegal trade globally, poses a serious threat to ecosystems and species diversity [1]. This study aims to address this pressing issue through a data-driven five-year program. Supported by hierarchical analytical modeling and random forest modeling, we developed a comprehensive plan for illegal wildlife trade with a focus on client needs. First, key factors were identified through hierarchical analytical modeling to determine potential customers. Next, multiple scenarios were analyzed using linear regression modeling to develop a five-year plan that meets customer needs. Subsequently, a random forest model was applied to validate the rationality of the program, and the predictive performance was continuously refined through successive trainings to provide the client with recommendations for additional resources and permissions, and to assess the impact and probability of success of the program. Ultimately, the combination of the AHP model and the random forest model indicates that the program is expected to significantly reduce illegal wildlife trade and make a positive contribution to the protection of rare wildlife, the maintenance of ecological balance and human well-being.

2. ANALYTIC HIERARCHY PROCESS

Analytic Hierarchy Process (AHP) refers to the decision method that decomposes the elements which related to decision into levels such as goals, criteria, and schemes, and carries out qualitative and relevant analysis on this basis [2]. This method is a hierarchical weight decision analysis method proposed by Professor Satie [3].

2.1. Data Analysis Process

In order to determine the research objects of the five-year project, we use illegal wildlife trade as a key word to search in international governments. In order to make our result stand out, we draw a word cloud image about the frequency of different words, the cloud image in shown in Figure 1.

Figure 1. Word cloud image

2.2. Solution and Result

From the result we can see the America relevant search are more obvious, so in this contest we choose the U.S. Environmental Protection Agency to serve as our main customer. In order to study the America and relevant index, we use some periodical research websites such as CNKI, PubMed and Google Scholar to search for illegal wildlife, and then conducted a secondary search with the U.S. as the key word. Finally, the 16 indicators with the highest frequency in the U.S. were obtained as the index evaluation system of this problem. Then we sort out those evaluation index, we divide them into 3 first index: authority, interest and resource. The picture is shown in Figure 2 below.

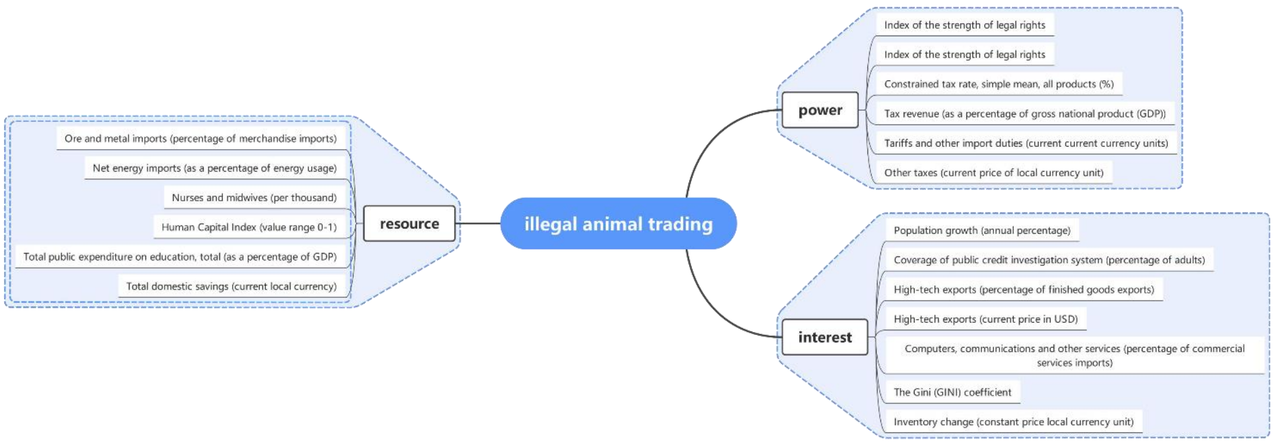

Figure 2. Evaluation index

Then we need to do the data cleaning procedure, we use Kolmogorov-Smirnov testing to decide our project's distribution mode [4], the results are shown in Figure 3 below.

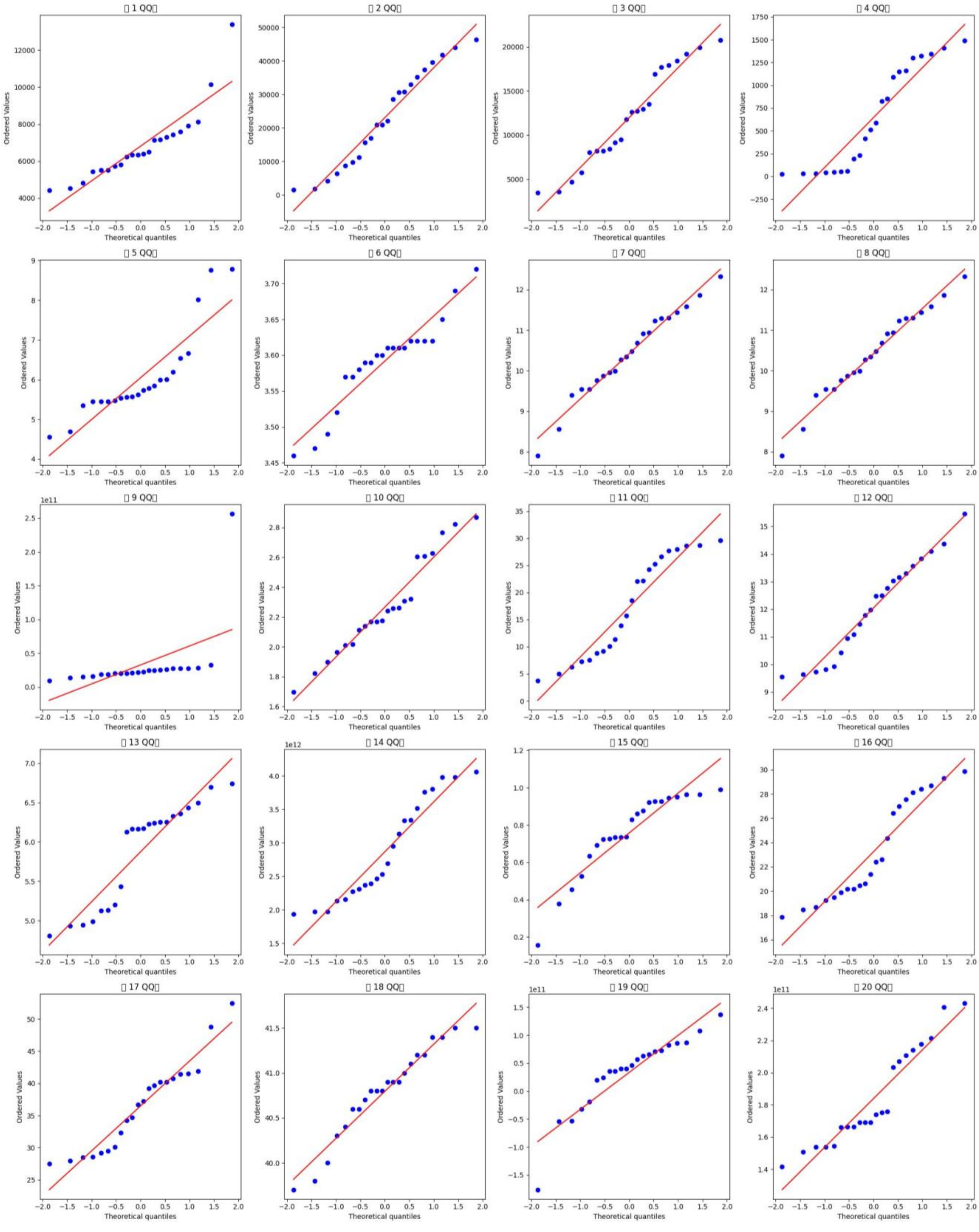

Figure 3. Kolmogorov-Smirnov test results

From Kolmogorov-Smirnov testing we are easy to find out the 1-9 group of data and the 11 groups of data are all conform to normal distribution.

We delete unnecessary data based on whether they are conforming to 3-sigma principal or not, next we are going to show one example based on the second group of data. The raw data is shown in Figure 4.

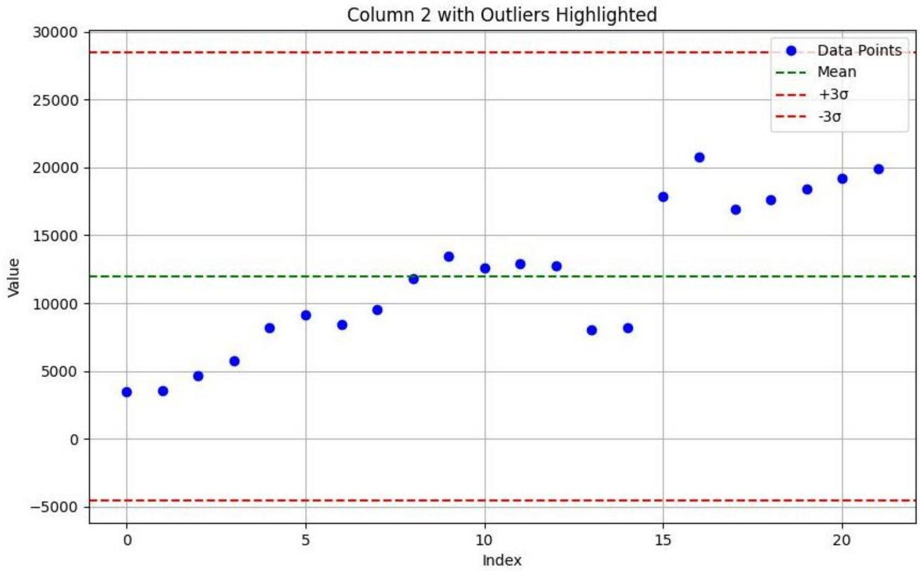

Figure 4. Before treatment

We then use Newton interpolation to make supplements for those deleted data, as shown in Figure 5.

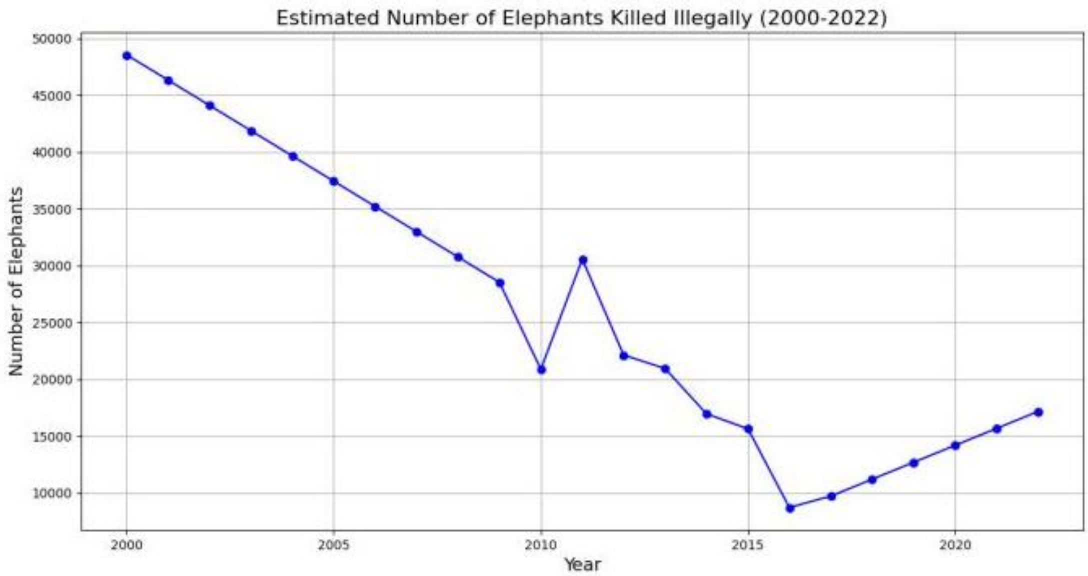

Figure 5. After treatment

At last, we make dimension reduction process to those figures and get our results in Figure 6.

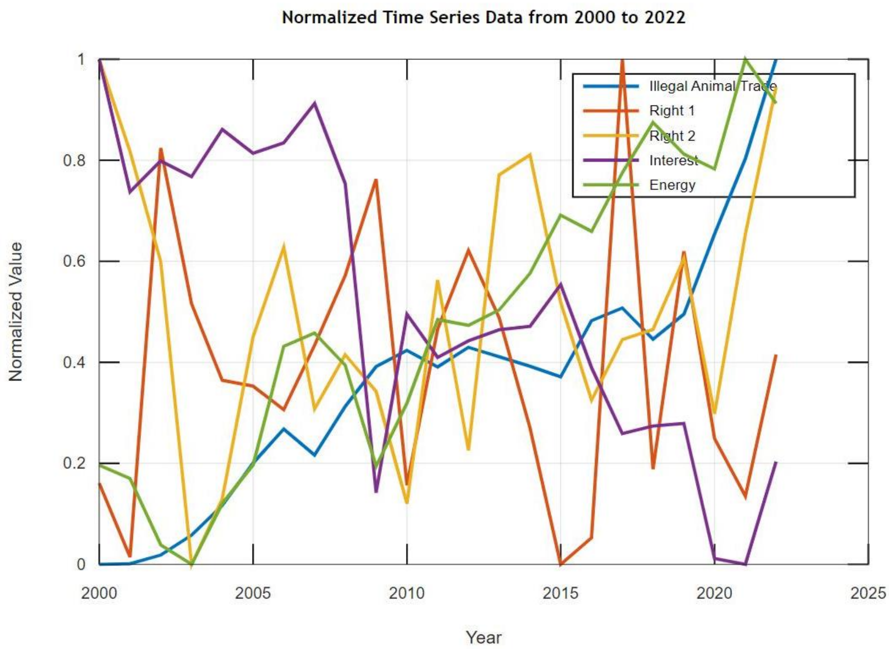

Figure 6. Normalized time series data from 2000 to 2022

3. LINEAR REGRESSION MODEL

A linear regression model is a mathematical regression model that determines the correlation between variables [5].

Three first-level indicators and four dependable variables selected in Section 2 were used to construct partial least square regression analysis, and the relationship between each dependent variable and the three indicators were analyzed. In case of 5 -year intervention, we directly used the linear regression model to predict the three independent variables. The predicted value without the intervention is compared with five-year project to illustrate the client. As for the solution suitability analysis: we use the U.S. Environmental Protection Agency as a client.

For the choice of power, resources and interests with the change of timeline graph, through the dimension reduction results of question one, we can see that those four things have very good connection with time. Hence, we establish a multiple linear regression model between the four indicators and the illegal wildlife trade.

We consider the illegal wildlife animal trade as variate Y, power1, power2, interest and energy to serve as independent variate X_1, X_2, X_3, X_4. Then we can find: $Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \varepsilon$. Among them, β_0, β_1, β_2, β_3, β_4 are coefficients, ε is error term.

The results are shown in Figure 7 below.

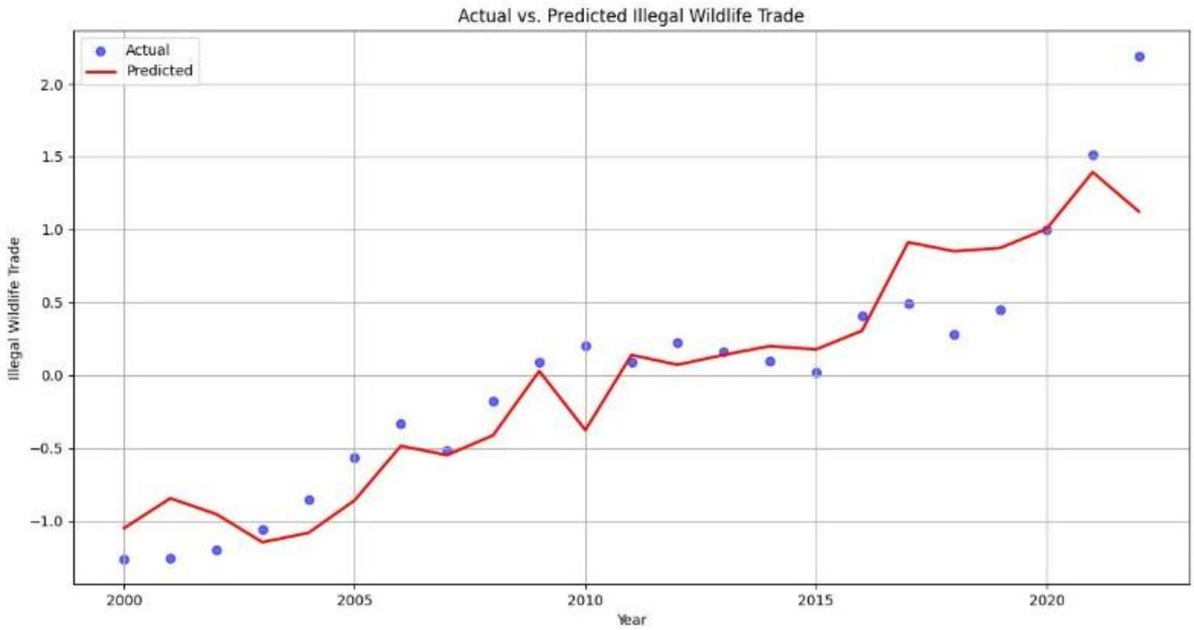

Figure 7. Actual vs. predicted illegal wildlife trade

Next, a unitary regression analysis is performed for these four indicators, as shown in Figure 8.

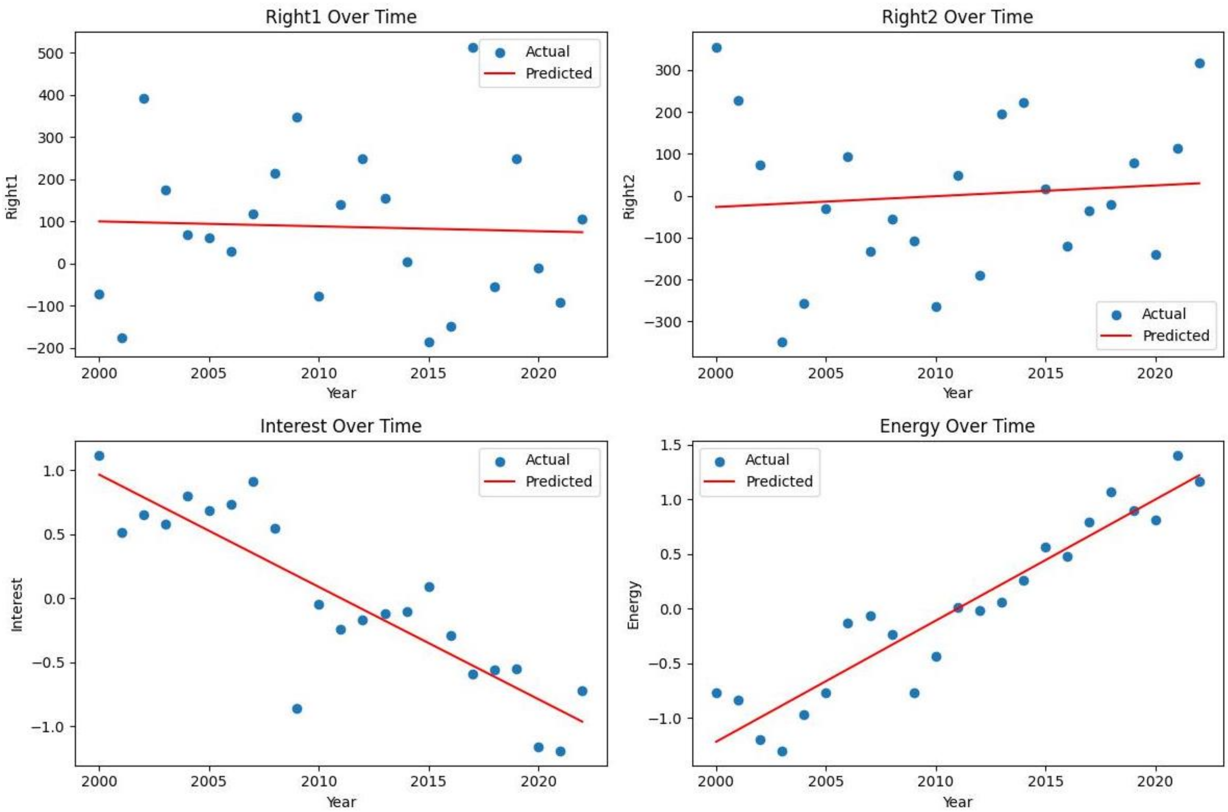

Figure 8. Regression analysis results

Then we make five years prediction for those unitary regression models, and we put the result into the multiple predictors model to get the result in Figure 9.

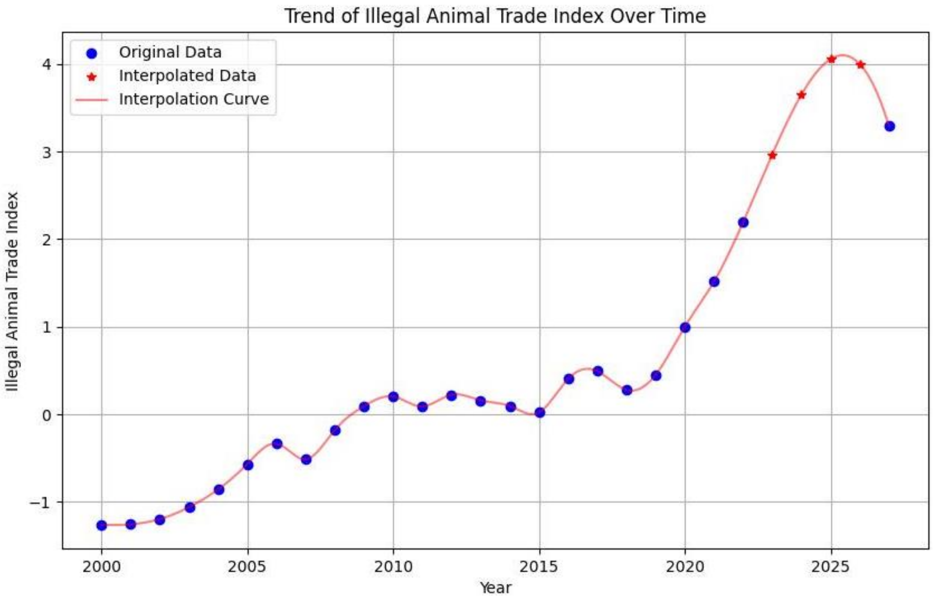

Figure 9. Regression result

Combined those figures, we can find that the implementation of the project has great significance in reducing illegal wildlife trade.

4. RADOM FOREST MODEL

In machine learning, a random forest is a classifier that contains multiple decision trees, and the category of its output is determined by the mode of the categories output by the individual trees [6]. The structure of random forest is shown in Figure 10.

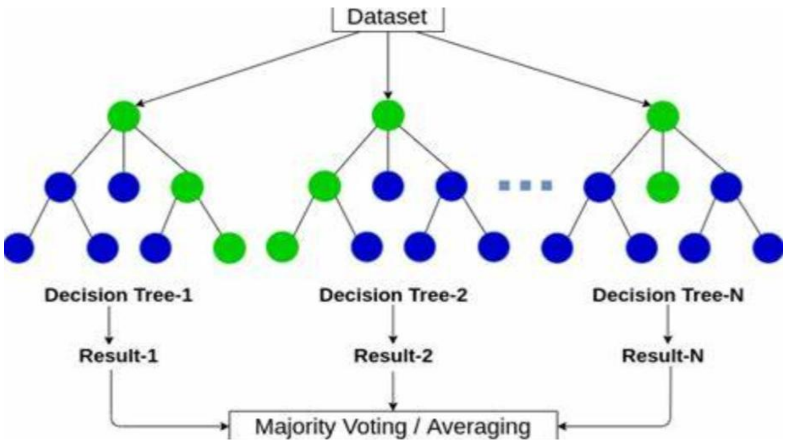

Figure 10. Random forest structure

4.1. Data Analysis Process

We need to first finish the data preparations, to start with, we have to choose the several features from the available data which is relevant with the illegal wildlife trade, for example: time, place, and species among the previous 16 features in question two.

Then, we need to finish the data processing procedure. To be more specific, there are two important steps: missing value processing and the data normalization. For numerical features, normalization is performed so that model can learn better. The next step is data fractionation---we have to divide the data set into training sets and test sets for model training and evaluation. The last step is to use the Scikit-Learn library in python, then we build a random forest classifier by adjusting hyperparameters.

4.2. Solution and Result

After establishing the random forest model classifier, we use Matplotlib and seaborn to do the data visualization.

The specific visualize data photos are shown in Figure 11 below:

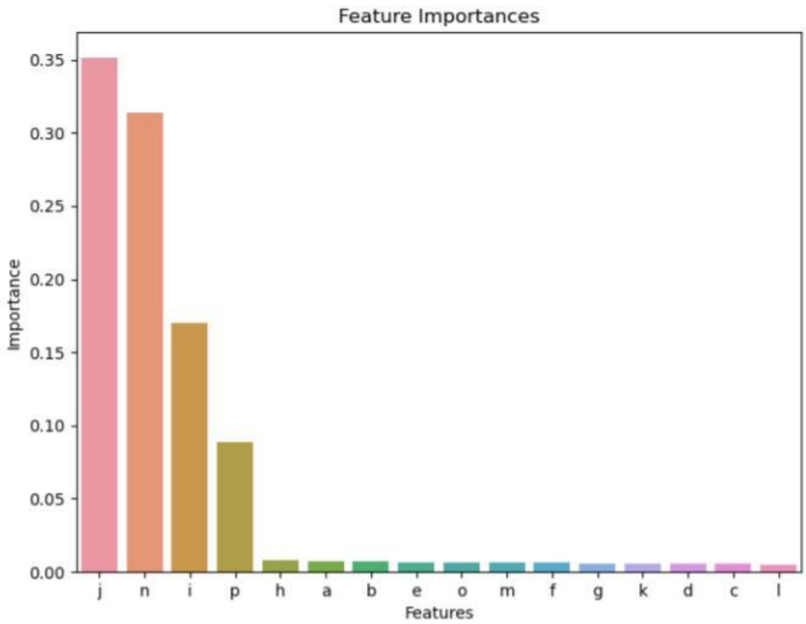

Figure 11. Visualization results

Among those 16 features, we found that the higher the importance is, the higher suitability it has with the project, based on this knowledge we can rank those features and find the additional power and resources that clients need.

5. THE ESTABLISHMENT OF EVALUATION MODEL

5.1. Analysis Process

The model predicts that the introduction of drones will prevent poaching by 30 to 50 percent, and this percentage will increase further if more drones are added, but the corresponding cost will also increase. In order to enhance the anti-jamming capability of the model and avoid reducing its usefulness lower down due to the high sensibility of the background, we take the following measurements:

We add sufficient noise to the possessing of feature values. Since we only have an average of poaching data over the years, we added a larger noise value of 0 to 0.4, but it didn't affect the average of the data.

We added many random values, such as the number of random steps in a random walk, the random area selection of the drone, and the random addition of noise.

Finally, we use r2 determination coefficient to measure the accuracy of our model. According to the operation results, the accuracy of the training data reached more than 90%. The accuracy of the random forest model is quite high and it's very reliable.

5.2. Result

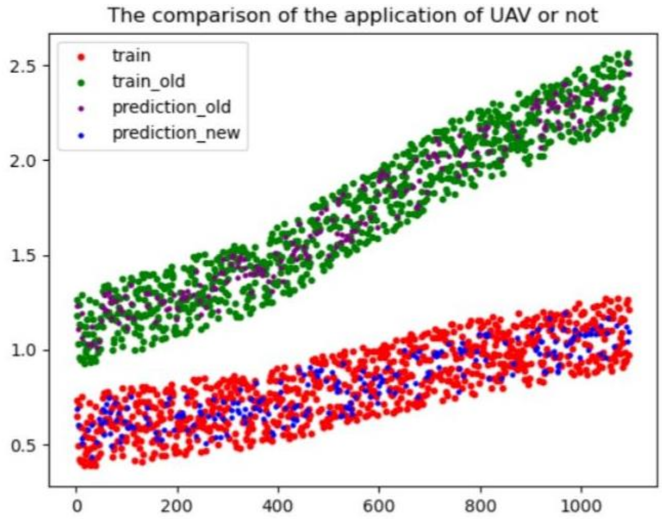

Figure 12. Predicted results

As shown in Figure 12, the top band is the poaching condition before the use of drones, and the bottom band is the poaching condition after the use of drones, which dramatically reduced daily poaching.

6. CONCLUSIONS

This study is based on a data-driven five-year program designed to address the growing problem of illegal wildlife trade. Supported by hierarchical analytical modeling and random forest modeling, we successfully identified key factors and identified potential clients to develop a five-year plan that meets their needs. Our linear regression model analyzed multiple scenarios and provided a solid foundation for the project. Validation of the Random Forest model further ensured the soundness of the project, and continuous training improved predictive performance. By integrating the AHP model and the random forest model, we conclude that this project is expected to significantly reduce illegal wildlife trade and make a positive contribution to the conservation of biodiversity and ecological balance. The integrative and prospective nature of this research methodology provides a new approach and practical path to address global environmental challenges. We look forward to putting this program into practice to save precious wildlife and protect our planet and human future.

REFERENCES

[1] Margulies J D, Bullough L A, Hinsley A, et al. Illegal wildlife trade and the persistence of “plant blindness” [J]. Plants, People, Planet, 2019, 1(3): 173-182.

[2] Subramanian N, Ramanathan R. A review of applications of Analytic Hierarchy Process in operations management[J]. International Journal of Production Economics, 2012, 138(2): 215-241.

[3] Wang L. Performance Evaluation and Management Method of Music Tutoring Schools Based on Analytic Hierarchy Process[J]. Discrete Dynamics in Nature and Society, 2022, 2022.

[4] Aslam M. Introducing Kolmogorov–Smirnov tests under uncertainty: an application to radioactive data[J]. ACS omega, 2019, 5(1): 914-917.

[5] Maulud D, Abdulazeez A M. A review on linear regression comprehensive in machine learning[J]. Journal of Applied Science and Technology Trends, 2020, 1(2): 140-147.

[6] Denisko D, Hoffman M M. Classification and interaction in random forests[J]. Proceedings of the National Academy of Sciences, 2018, 115(8): 1690-1692.