LIFE: International Journal of Health and Life-Sciences
ISSN 2454-5872

Dave et al., 2021

Volume 7, pp. 123-130

Received: 29th September 2021

Revised: 3rd November 2021, 9th November 2021, 11th November 2021

Accepted: 14th November 2021

Date of Publication: 17th November 2021

DOI- https://dx.doi.org/10.20319/lijhls.2021.7.123130

This paper can be cited as: Dave, K., Barber, S. & Verma, P. (2021). A Machine Learning Linear Regression Model to Predict Future Giant Panda Population. LIFE: International Journal of Health and Life-Sciences, 7, 123-130.

This work is licensed under the Creative Commons Attribution-Noncommercial 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

A MACHINE LEARNING LINEAR REGRESSION MODEL TO PREDICT FUTURE GIANT PANDA POPULATION

Kumud Dave
Student (B.Sc.), Guru Nanak Girls College, Udaipur, Rajasthan, India
kumuddave2002@gmail.com

Satyendra Barber
Assistant Professor, Guru Nanak Girls College, Udaipur, Rajasthan, India
barbersatyen@gmail.com

Pooja Verma
Assistant Professor, Guru Nanak Girls College, Udaipur, Rajasthan, India
dr.poojaverm21@gmail.com

Abstract

Increasingly used as the insignia of China, the zaftig and enchanting Giant Panda lives on mountains of southwest China. The Giant Panda is on the WWF logo and is known as "National Treasure" in China. In this study, we predict the future Giant Panda population by using machine learning algorithms of the simple linear regression model. We take different variables to predict the next 30 years of the Giant Panda population. Focusing on the factors which affect the Giant Panda population. We take several parameters for this research like Bamboo Population, Annual Rainfall in China, Carbon Stock in Bamboo Stems, Deforestation, and Human Influence and Population of Giant Panda. Despite their peak status and relative deficiency of natural predators, pandas are still at risk and multiple intimidations from human

influence have left just over 1,800 Pandas in the forest. To be ready for future troubles it is mandatory to have a pre-look of some conditions so that we can be prepared for that. Substantially, Endangered species at the edge of extinction are kept in extra special conservation. The machine learning algorithms developed with a wide-ranging of training datasets that help to find results faster and accurately.

Keywords

Conservation, Giant Panda, Machine Learning, Linear Regression, Deforestation

1. Introduction

The giant panda is also known as the panda bear and its scientific name is Ailuropoda melanoleuca. Unparalleled habitat makes giant pandas an illustration of umbrella species. By this speciality, they can protect many other species. If we protect the giant panda, we indirectly protect many other wild animals that survive around them like multi-coloured pheasants, the golden monkey, takin, and created ibis. Nowadays China is putting much more effort into saving these lovely creatures. After 1949, the giant panda population created pressure on the panda's habitat and because of the excessive famines, hunting of wild animals including pandas increased and because of the increasing demand of the giant panda's skin in many countries like Hong Kong and Japan poaching of the giant panda is escalating. By the 2006 report, the panda's population in China was only 1,000. The giant panda is the world's most admired and shielded rare animal and they are very few and rare, so they were able to gain an inhabitant status in UNESCO (United Nations Educational, Scientific and Cultural Organisation) world heritage site designation. Artificial intelligence (AI) is one of the major areas in computer science that can be used to solve several real-life problems. The main reason behind the decreasing panda population is habitat destruction. Algorithms that are used for various predictions are a scientific formula for establishing a relationship between the historical datasets. The development efforts for the growing China population are leading to habitat destruction. Panda gets the most threatening diseases like sunstroke (excessive exposure to the sun), ileus (obstruction of the ileum), volvulus (twisting of the loop of intestine), acute pancreatitis (inflammation of the intestine), and kidney failure.

2. Literature Review

Conceptual modelling is applied for hydrological forecasting but unfortunately, we cannot use this forecasting system for population estimation because the calibration data and

the population forecasting need advanced numerical tools. So, we use the linear regression model technique of supervised learning. In supervised learning, we are given a data set and receive output that we have already known. Linear regression is a statistical model that explores the linear relationship between two (simple linear regression) or more (multiple linear regression) variables. Nowadays modern researchers also use machine learning to predict modelling approaches and parameters to improve accuracy and precision of the model. Therefore, this study was conducted to predict the future population of the giant panda by which we will become aware and prepared for future challenges.

3. Problem Statement

Giant pandas became endangered in 1990 because of excessive poaching and deforestation. Expansion of farm development causes fragmentation of panda habitat and fragmentation leads greatest threats to the continued remain of the species. A large area of the giant panda's habitat has already been extinct for timber and fuelwood. Other areas have been cleaned for agriculture. So, there is a need to develop a machine learning model that is capable of predicting population estimates with good accuracy and precision. A Machine Learning model is used in predicting the population of giant pandas based on historical data. An accurate ML model could give an early alert of potential endangered conditions of Panda.

4. Objectives

The major aim of the study is to develop a strong model for predicting the population of the giant panda and save them from future trouble. This study is vital to enhancing the population of giant pandas in China.

5. Case Study and Data Description

Data for this study consist of several parameters like the population of the giant panda, bamboo population (area million hectare), annual rainfall in China(mm), carbon stock in bamboo stems (tg c), deforestation(hectare), human influence (China population) etc. In this study, we first visualize the data and process it further to build a simple linear regression model. In figure 1 we can see that annual rainfall in China increases with a decrease in giant pandas because an excessive amount of rainfall causes many diseases and harm to the giant panda. Similarly, we have a total of 6 independent variables that are used to train and build our regression model. In figure 2 we can see that carbon stock in bamboo stems also increases with

the decrease in the population of the giant panda. Carbon stock in bamboo stems and bamboo population increases because annual rainfall increases but excessive rainfall causes many seasonal diseases in the giant panda. In figure 3 the population of giant pandas gradually increases with increase in bamboo population. In figure 4 deforestation is continuously decreasing. In figure 5 human influence increase.

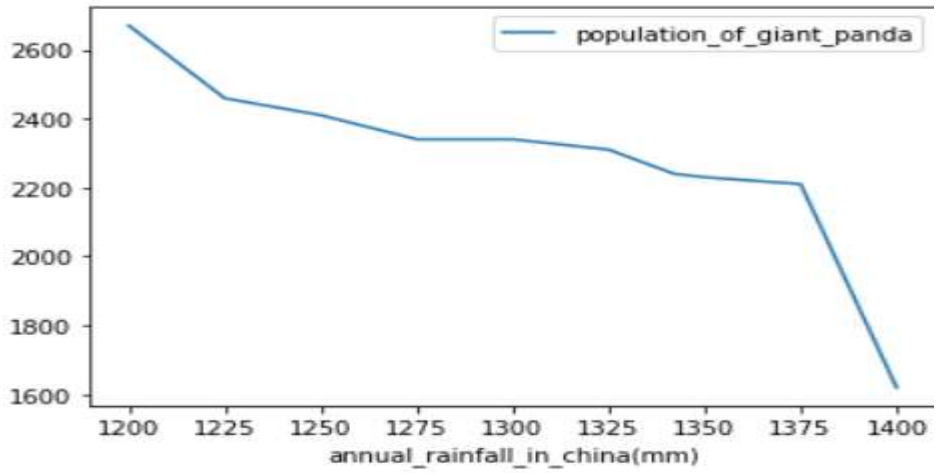

Figure 1: *Graph Between Annual Rainfall in China Vs Population of Giant Panda*
(Source: www.researchgate.net)

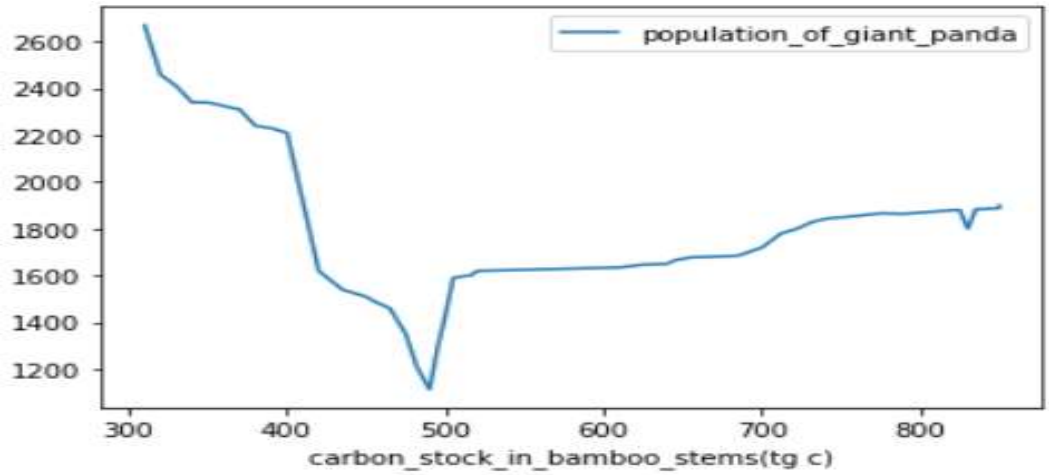

Figure 2: *Carbon Stock in Bamboo Stems Vs Giant Panda Population*
(Source: Self)

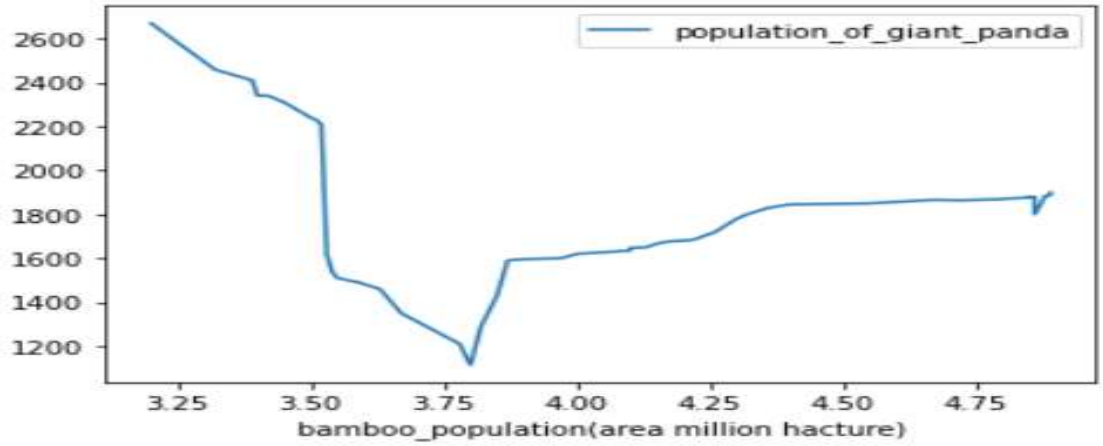

Figure 3: *Bamboo Population (Area Million Hectare) Vs Giant Panda Population (Source: Self)*

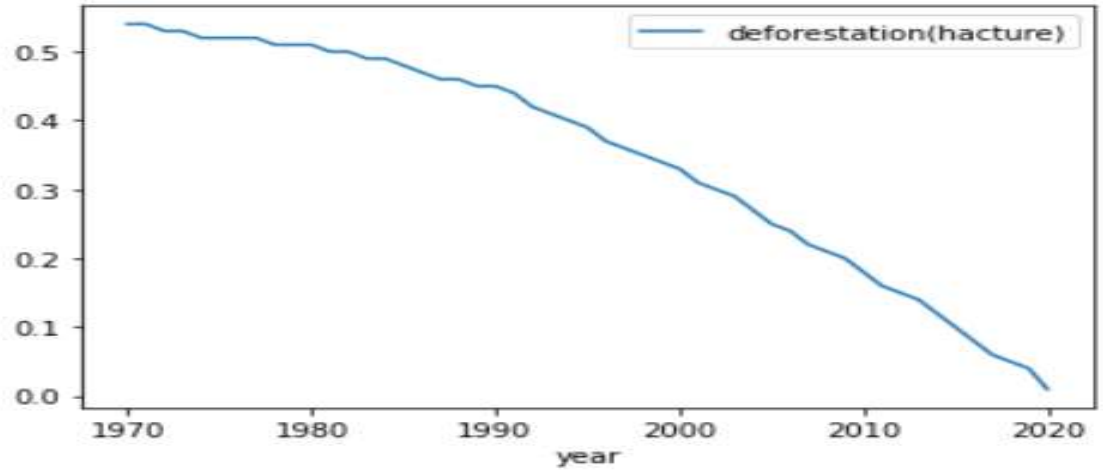

Figure 4: Deforestation Vs Time
(Source: Self)

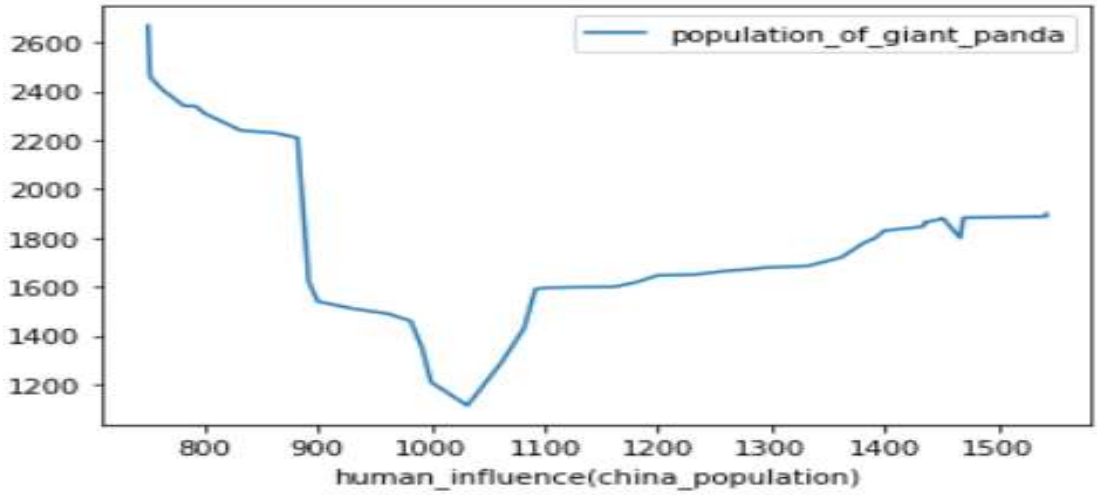

Figure 5: *Human Influence (China Population) Vs Population of Giant Panda (Source: Self)*

6. The Model Used for Forecasting

In this study, we use a multiple linear regression model. Multiple linear regression creates the relationship between 2 or more variables and is a statistical technique that uses many explanatory variables to predict the outcome of the dependent variable. the relationship between the dependent variables and explanatory variables is presented by the following equation:

6.1. For Simple Linear Regression

$$y = b0 + b1*x1$$

Where, x1= independent variable

Y= dependent variable

b0=intercept

b1= slope coefficient

6.2. For Multiple Linear Regression

$$y = b0 + b1*x1+b2*x2+\ldots. +bn*xn$$

Formula to get b1 and b0.

$$b1=\frac{n\sum xiyi-\sum xi\sum yi}{n\sum xi^2-(\sum xi)^2}$$

$$b0 = \overline{y} - b1\overline{x}$$

R-squared formula: -

$$r = \frac{n\,(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2]\,[n\sum y^2-(\sum y)^2]}}$$

where,

r = The correlation coefficient

n = Number in the given datasets

x = independent variable

y = dependent variable

Using the train test split algorithm, we classified the training dataset and testing dataset.

7. Methodology

We developed a machine learning multiple linear regression model with actual giant panda population data from the year 1970 to 2020 and extrapolated to predict population

growth till 2050. Google Collab and MATLAB machine learning tool used for this research graphs. We draw the scatter plot between dependent and independent variables by python scripting and train the data to get a regression equation and its accuracy is defined by the R-squared value.

8. Result and Discussion

Our goal of this work was to predict the next 30 years of the giant panda population by using the trained model.

Table 1: *Data of Predicted Giant Panda Population*

Year	2025	2030	2035	2040	2045	2050
Bamboo Population (area million hectare)	7.8	8	8.6	10	11.3	12
Carbon stock in bamboo stems (tg c)	1050	1210	1350	1420	1500	1600
Deforestation(hectare)	0.18	0.16	0.15	0.14	0.12	0.1
Human Influence (China population)	1346	1400	1385	1359	1340	1310
Annual Rainfall in China(mm)	580	585	590	575	600	607
Predicted Giant Panda Population	2870	2248	2168	3222	4192	4480

(Source: Self)

Based on independent parameter values gathered from different research papers from various journals, we predict the population count of 2025 will be 2870 with 96% R-squared value.

Table 2: *Regression Results*

OLS Regression Results

Dep. Variable:	population_of_giant_panda	R-squared (uncentered):	0.967
Model:	OLS	Adj. R-squared (uncentered):	0.966
Method:	Least Squares	F-statistic:	1470.
Date:	Wed, 02 Jun 2021	Prob (F-statistic):	9.75e-39
Time:	19:40:44	Log-Likelihood:	-368.07
No. Observations:	51	AIC:	738.1
Df Residuals:	50	BIC:	740.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
annual_rainfall_in_china(mm)	1.3693	0.036	38.334	0.000	1.298	1.441

Omnibus:	8.811	Durbin-Watson:	0.138
Prob(Omnibus):	0.012	Jarque-Bera (JB):	8.018
Skew:	0.863	Prob(JB):	0.0182
Kurtosis:	3.893	Cond. No.	1.00

(Source: Self)

With 96% accuracy there will be near 4192 population of giant panda in 2045 and nearly 4480 population of giant panda in 2050. This trained Model performs well on the task of projecting and predicting population estimates.

REFERENCES

Buchholz, K. (2019). Panda Population in the Wild Rebound. PANDA CONSERVATION. https://www.statista.com/chart/18740/panda-populatins-in-the-wild-and-in-captivity/

Lucas Gutierrez Rodriguez, (2014). From basic raw material goods to cultural and environmental services: the Chinese bamboo sophistication path. Ecology and Society 19(4): 3. https://www.researchgate.net/figure/Evolution-of-bamboo-total-area-in-china-National-Bureau-of-Statistics-of-china_fig1_267641764

Muhammad Tauseef, Lihua Chen, Kaipeng Yang, Yunyao Chen, (2020). "Long-Term Rainfall Trends and Future Projections over Xijiang River Basin, China", Advances in Meteorology, vol. 2020, Article ID 6852148, 18 pages https://doi.org/10.1155/2020/6852148

Ritchie, H. & Rosser, M. (2021). "Forest and Deforestation". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/forests-and-deforestation' [Online Resource] https://ourworldindata.org/deforestation

Thang Muoi mot, (2019). Line chart 3: Population growth and projection.http://fastlearn.edu.vn/2019/11/25/line-chart-3-population-growth-and-projection/

Xinzhuang Chen, G.U.O.M.O... Z.H.O.U. (2020). Changes of carbon stocks in bamboo stands in China during 100 years. Forest ecology and management 258(7):10. https://www.sciencedirect.com/science/article/abs/pii/SO378112709004745

Zhao, J. & Wang, Z. (2020). Future trend of water resources and influences in agriculture in China. https://www.researchgate.net/publication/340725955_Future_trends_of_water_resources_and_influences_on_agriculture_in_China