Asian Journal of Economics, Business and Accounting

Volume 24, Issue 3, Page 221-228, 2024; Article no.AJEBA.112671
ISSN: 2456-639X

The Prediction of Growth of the GDP of North Macedonia for 2024 using Logistic and Linear Regression Model

Leotrim Ramadani [a], Neritan Turkeshi [b] and Shpëtim Rexhepi [a*]

[a] Faculty of Informatic Sciences, Mother Teresa University, Skopje, North Macedonia.
[b] Faculty of Technical Sciences, Mother Teresa University, Skopje, North Macedonia.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJEBA/2024/v24i31255

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/112671

Original Research Article

Received: 08/12/2023
Accepted: 12/02/2024
Published: 16/02/2024

ABSTRACT

The purpose of this research is to get to know the statistical calculations that are used even in other fields except Mathematics, like economics, Finance, Informatics, Architecture, etc.
In this research, we have used statistical calculations like those of Linear Regression and Logistic Regression to calculate and predict the GDP of a specific country, in our case, North Macedonia. Here we have used the chance to use these two statistical models to present and visualize the already made predictions. The data are derived from the World Organization which deals with statistical data of all the countries of the world.
While using these two statistical models we even explained which are the pros and cons of each other, when we should use the first one and when the second one, in what use-cases we can use these two and what are the differences between them.

**Corresponding author: E-mail: Shpetim.Rexhepi@unt.edu.mk;*

Keywords: Logistic; linear regressions; GDP; North Macedonia.

1. INTRODUCTION

National income and product accounts—best known by one of their principal aggregates, gross domestic product (GDP)—are produced by virtually every nation in the world [1]. During recessions, upward-sloping yield curves not only indicate bad times today but better times tomorrow. Guided by this intuition, many papers predict GDP growth in OLS regressions with the slope of the yield curve, usually measured as the difference between the longest yield in the dataset and the shortest maturity yield [2].

In this research, we tried to predict the GDP of North Macedonia for 2024 with the help of Logistic and Linear research. Here we even have explained what exactly these two functions are. We have used these two to calculate and visualize the results and the data using the programming language called Python.

The GDP of a country is important to know because it informs us of the economic situation of a country and how it works. With the usage of Logistic regression, we get discrete data and information on whether we will have an increase in the GDP or a decrease.

Linear Regression on the other hand gives us more concrete and it tells us more “precisely” the percentage of the increase or the decrease of the GDP of a specific country.

In this research, we will get to know what is the GDP of a country, what it is needed for, and how it is calculated.

As an important point of this research, we also have the explanation of how to calculate the GDP with the use of some Math functions and how to visualize it with the connection between Math, Economy, and Informatics or programming.

With the connection of each other we can achieve many great things because IT as a science cannot exist without Math functions and Math or its’ functions and fields, we could not visualize without the help of IT science.

This was the main purpose of this research and in the end, we achieved this connection of course we have explained how we did it and we have shown its results.

1.1 Similar Examples Like our Research

Linear regression to calculate the diabetes of a specific person. In the example of the Linear Regression, we have shown the code and the explanation of how the prediction is made and what result it returns. [3] (*Linear Regression Example*, o. J.). This blog has helped us to better understand the algorithm of linear regression and how we can adapt it to the purpose of our research.

In this blog [4] (*Python Machine Learning Linear Regression*, o. J.-a) we also have some good posts where we can learn how to code a prediction and how to learn about linear regression. W3schools is a blog that offers us more opportunities to learn more things in a lot of programming languages. To learn more about Linear Regression in the programming language called Python, we can use this blog [5] (Python, o. J.) to read more about it theoretically. Here we have a much better explanation of the regression and its algorithm.

- Examples of calculating the GDP of a country we have in this blog [6] (*Measuring the Size of the Economy*, o. J.) which was written by Khan Academy e and published in 2016.

1.2 Logistic Regression and Linear Regression

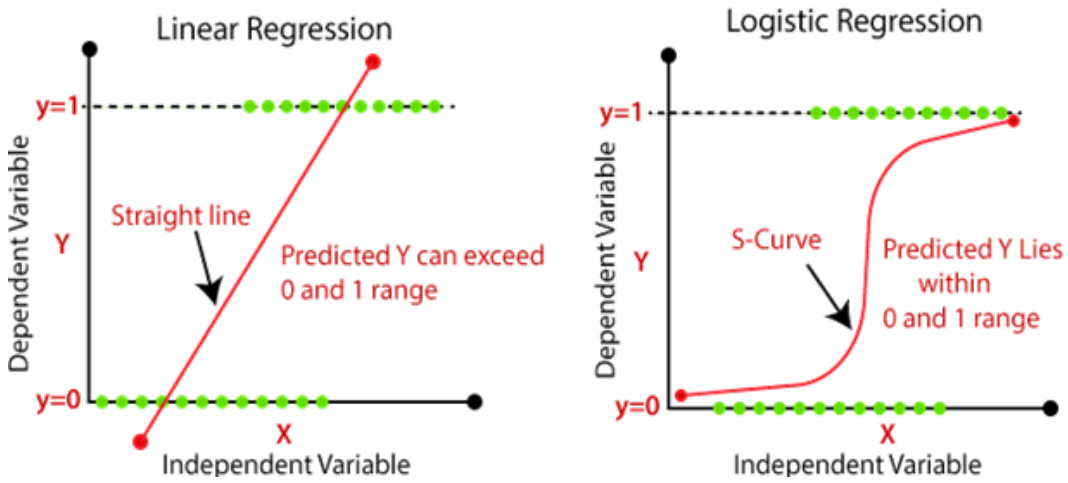
Logistic and linear regression are two kinds of statistic models that are used in machine learning but both of them have different purposes and are applied in different fields.

Being different models that are both from supervision learning, the only thing that distinguishes them is where are they applied.

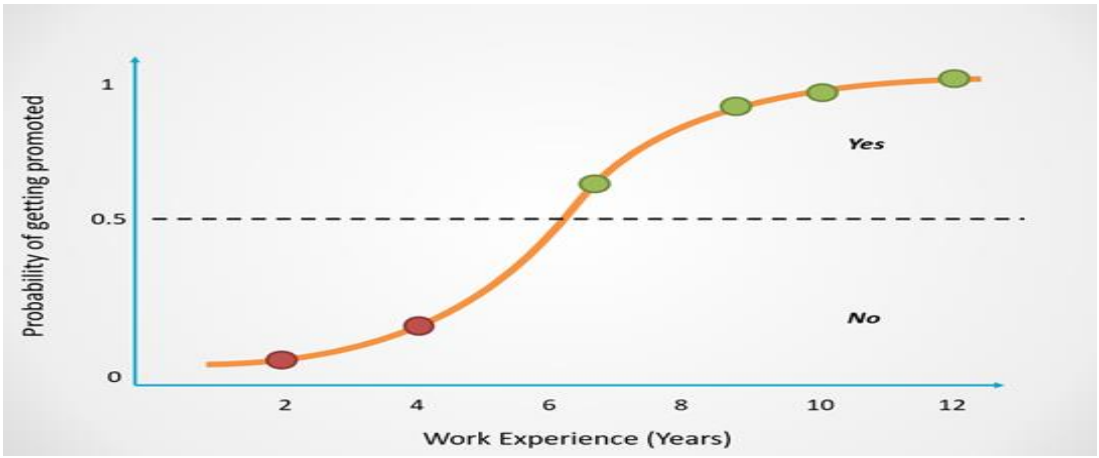
Linear regression is used for problem solving of the regression and logistic regression is used for solving problems that are about the data classification.

Simple linear regression is applied to estimate the relationship between the dependent variable, y, and the single explanatory variable, x, by taking a set of data that includes observations of these two variables for a given population [7].

the picture 1 shown image we see an example of all this:



Picture 1. Linear and logistic regression model



Picture 2. Algorithms in machine learning applications

1.3 Logistic Regression

Logistic Regression is a process of modelling the probability of a discreet result.The most common case of a Logistic Regress returns a binary result, i.e.: true/false, 0/1, Yes/No etc. Binary variables are used widely in statistics to create probability models for a specific class or a specific scenario, for example: the probability for a football club to win a match against another club, probability to predict the health of a patient, for different calculations in Economy and other different fields.

Logistic Regression nowadays has become a very important tool in the discipline of the machine learning. This allows the used algorithms in machine learning applications to classify the input data based on previous inputted data or even historical data. With the input of the additional input data, these algorithms get improved on predicting the classification between these data.

Example 1:

This picture 2 makes the explanation of the Logistic very easy! In the X axis we have shown the years of experience of one or more employees and in the Y axis we have the probability that tells us which employee with how much years of experience is more likely to get promoted to a higher rank or get a higher salary.

We can easily tell that depending on how many years of experience an employee has, his probability to get promoted gets higher and higher.

This is a very simple and understandable example of the logistic regression.

Example #2:

This picture 3 example also shows a very nice and simple concept on logistic regression.

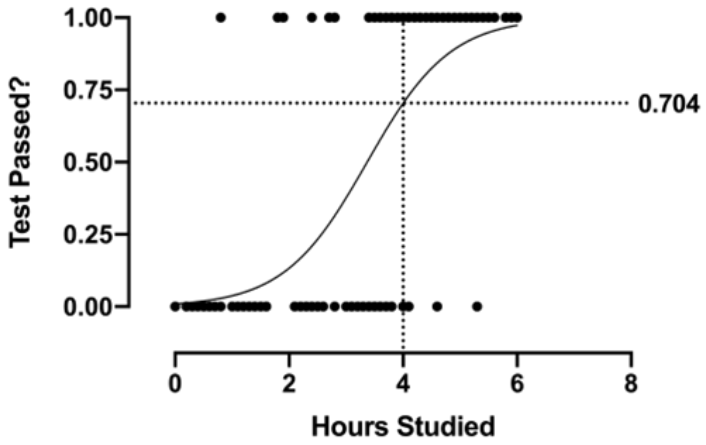
Here we can see that if a student spends more time studying then his probability to pass the class gets always higher and higher.

The line in the middle of the graph is printed based on the points or the data that are collected from the student in this use-case. So, if a student spends every day more and more hours studying this line will go to 1, which means the student will surely pass. On the other hand, if the student spends one week studying and then takes a break for two weeks and then comes back to study, this line will be almost on the lowest point and of course the probability for the student to pass the class will be very low.

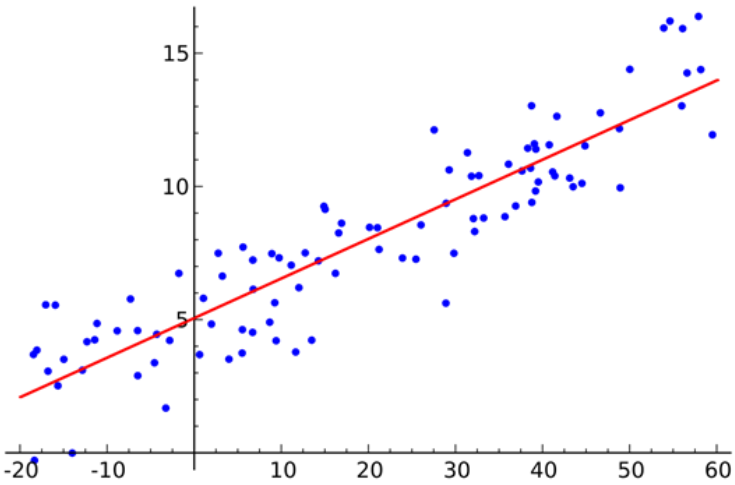
1.4 Linear Regression

Linear regression analysis is used to predict the value of one variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you use to predict the value of another variable is called the independent variable.

Linear regression was the first type of regression analysis to be rigorously studied and widely used in practical applications. This is because models that depend linearly on their unknown parameters are easier to fit than models that are nonlinearly related to their parameters, and because the statistical properties of the resulting estimators are easier to be determined.



Picture 3. Understandable example of the logistic regression



Picture 4. Linear regression analysis

This form of analysis estimates the coefficients of the linear equation including one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least-squares" method to find the line of best fit for a set of paired data.

1.5 Advantages and disadvantages of Linear Regression

1.5.1 Advantages

- **Simple model:** The linear regression model is the simplest equation used to express the relationship between multiple predictor variables and the predicted variable.
- **Efficient calculations**: The speed of linear regression modeling is fast as it does not require complicated calculations and performs predictions quickly when the amount of data is large.

Interpretability of the result: The ability of linear regression to determine the relative influence of one

or more predictor variables on the predicted value when the predictors are independent of each other is one of the main reasons for the popularity of linear regression.

1.5.2 Disadvantages

- **Simplicity**: The linear regression model is too simple to capture the complexity of the real world.
- **The independence of the variables** It assumes that the predictor variables are uncorrelated, which is rarely true. Therefore, it is important to remove multicollinearity (using dimensionality reduction techniques), because the technique assumes that there is no relationship between the independent variables. In cases of high multicollinearity, two traits that are highly correlated will influence each other's weight and result in an unreliable model.
  • **Assumes homoscedasticity:** Linear regression looks at a relationship between the mean of the predictor/dependent variable and the predicted/independent variables and assumes constant variance around the mean, which is unrealistic in most cases.

1.6 GDP of a country

Gross domestic product (GDP) is the total monetary or market value of all finished goods and services produced within a country's borders in a given period of time. As a broad measure of gross domestic product, it functions as a comprehensive scorecard of a given country's economic health. [8] (Wyplosz, 2017)The fact that the GDP data is obtained from the declarations of natural and legal persons immediately raises the suspicion that individuals and companies may misrepresent and hide their finances from the tax authorities. Such undeclared or falsified income and calculations often show a poor country with a low GDP.The inaccuracy of these calculations or data is worrying because based on them, the governments of a country determine economic policy, investors evaluate their assets and companies decide to hire more workers or lay off workers and buy new plants and equipment.

Another risk that can appear in a certain country is the transition through the inflation phase.Inflation occurs when many daily products become more expensive and become unaffordable for the common population or for the middle and lower class.

Calculating a country's GDP includes all private and public consumption, government spending, investment, additions to private inventories, construction costs paid, and the external balance of trade. Exports are added to the value and imports are subtracted.

The formula for calculating the GDP: GDP=C+G+I+NX

C-Annual Consumption
G-Government expenses
I-Government investments
NX- Neto Exports

1.7 Predicting North Macedonia's GDP for 2024 with the help of Logistic Regression

As we mentioned above, this type of regression shows us only positive or negative values. Below we will have the coding of this prediction and we will be able to see what result we can get.

After testing the linear regression we reached this result. This means that the algorithm and regression used is 70% sure that in 2024 we will have an increase in GDP.

It is also worth noting that to arrive at this result, static and ordinary data were used, that is, they are not real or accurate.

1.8 Results from the prediction with the help of Logistic Regression

The moment that we execute the code that we use for these testing and predictions, then we automatically will win the shown result in the picture 5 below this text. In this picture we can easily see the result that we won from the prediction and what kind of data are generated to do the testing.

Here we have the result and the percentage of how accurate this prediction can be. For these static data, we obtained a positive result and that we will have an increase in GDP and that the percentage of accuracy is about 70%.

Since logistic regression uses fixed data and that it can only be 0 and 1, we have used encoding functions in order to obtain a data string with this type of values.

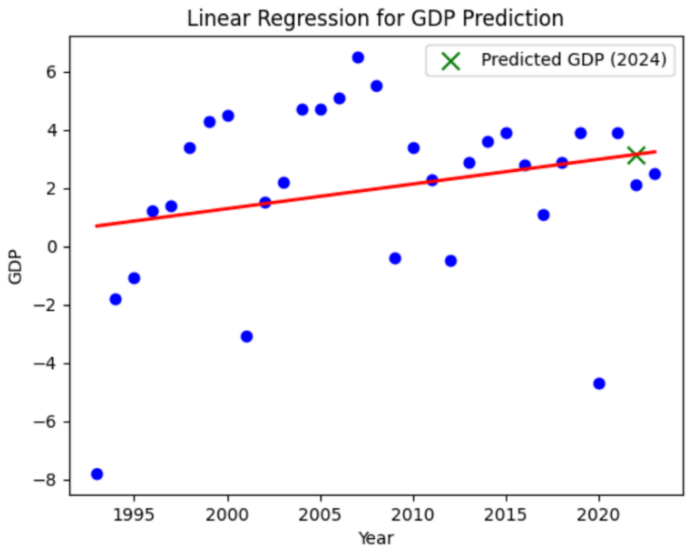
```
Accuracy: 0.70
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         5
           1       0.74      0.93      0.82        15

    accuracy                           0.70        20
   macro avg       0.37      0.47      0.41        20
weighted avg       0.55      0.70      0.62        20

y= 0     1
1     0
2     0
3     1
4     1
     ..
95    1
96    1
97    0
98    1
99    0
Name: GDPGrowth, Length: 100, dtype: int32
```

Picture 5. Accuracy classification report



Predicted GDP for 2024: 3.15

Picture 6. Linear regression for GDP prediction

1.9 The Code for Logistic Regression

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

np.random.seed(42)

data = {
   'InflationRate': np.random.uniform(0, 5, 100),
   'UnemploymentRate': np.random.uniform(5, 20, 100),
   'GDPGrowth': np.random.choice([0, 1], size=100, p=[0.3, 0.7])
}

df = pd.DataFrame(data)
X = df[['InflationRate', 'UnemploymentRate']]
y = df['GDPGrowth']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LogisticRegression(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

print(f"Accuracy: {accuracy:.2f}")
print("Classification Report:\n", report)
print("y=", df['GDPGrowth'])
```

1.10 Predicting North Macedonia's GDP for 2024 with the Help of Linear Regression

Linear Regression has it's differences from the Logistic regression. It precisely shows the percentage of how much the GDP will be increasing or decreasing in a year.

The data are written in a X and Y axis and depending of the data, there will be formed a line which will show where are the most common predicted values. If the predicted value touches the line or is above it, that tells us that the GDP will become higher next year, but if the predicted point is below the line, then the GDP will be lower next year.

Below this text we have the results from our prediction which we have done with the help of linear regression.

1.11 Results from the Prediction with the Help of Linear Regression

Unlike logistic regression that uses static and primary data such as 0 and 1, linear regression on the other hand uses dynamic data and must be well located and accurate, otherwise we will have inaccurate prediction.

We have received these data from a World Organization that deals with statistical data of all the countries of the world. Here we could also find data for North Macedonia and we could download them from this link: https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/MKD .

When we executed the code we got the results that are shown in the picture 6 this text and there we can see that North Macedonia is having success on increasing its' GDP. In that picture we also have the points in which NMK has had its' own crisis, i.e.: in 1993 where the GDP was down to -7.8% and in the 2020 where because of the global pandemic, NMK's GDP was down to -4.7%.

1.12 The Code of Linear Regression

```
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
years = np.array([1993,1994,1995,1996,1997,1998,1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021,2022,2023])
gdp = np.array([-7.8,-1.8,-1.1,1.2,1.4,3.4,4.3,4.5,-3.1,1.5,2.2,4.7,4.7,5.1,6.5,5.5,-0.4,3.4,2.3,-0.5,2.9,3.6,3.9,2.8,1.1,2.9,3.9,-4.7,3.9,2.1,2.5])
years = years.reshape(-1, 1)
model = LinearRegression()
model.fit(years, gdp)
next_year = np.array([[2022]])
predicted_gdp = model.predict(next_year)
plt.scatter(years, gdp, color='blue')
plt.plot(years, model.predict(years), color='red', linewidth=2)
plt.scatter(next_year, predicted_gdp, color='green', marker='x', s=100, label='Predicted GDP (2024)')
plt.xlabel('Year')
plt.ylabel('GDP')
plt.title('Linear Regression for GDP Prediction')
plt.legend()
```

```
plt.show()
print(f'Predicted GDP for 2024: {predicted_gdp[0]:.2f}')
```

2. CONCLUSION

In this research we learned about calculating the GDP of a country, it's formula, about the connection between Math, Economics and Informatic sciences just so we can use linear and logistic regression to calculate the GDP of North Macedonia.

Based on the data that we generated automatically via some code and based on the real data that we have collected from World Organizations for statistics, we discovered that the GDP of our country is pretty stable and it has the capacity for even larger increases.

In both two predictions that we made, we saw that North Macedonia will have an increasing GDP for 2024 and that the accuracy for that is around 70%.

In the first prediction we only saw whether we will have an increase or a decrease of the GDP, but in the second prediction we saw that in 2024 we will have an GDP percentage that goes up to 3.15%.

The usage of Informatics sciences with its' fields and with the science of Math proved us one more time that how important they are for each other, what do they serve for and what we can achieve from them.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Landefeld JS, Seskin EP, Fraumeni BM. Taking the pulse of the economy: Measuring GDP. Journal of Economic Perspectives. 2008;22(2):193-216.
2. Ang A, Piazzesi M, Wei M. What does the yield curve tell us about GDP growth? Journal of econometrics. 2006;131(1-2):359-403.
3. Linear Regression Example. (o.J.). Scikit-Learn; 10 December 2023. Available:https://scikit-learn/stable/auto_examples/linear_model/plot_ols.html
4. Python Machine Learning Linear Regression. (o.J.-a); 10 December 2023. Available:https://www.w3schools.com/python/python_ml_linear_regression.asp
5. Python R. (o.J.). Linear regression in python – Real python; 10 December 2023. Available:https://realpython.com/linear-regression-in-python/
6. Measuring the size of the economy: Gross domestic product (article). (o.J.). Khan Academy; 10 December 2023. Available:https://www.khanacademy.org/economics-finance-domain/macroeconomics/macro-economic-indicators-and-the-business-cycle/macro-the-circular-flow-and-gdp/a/measuring-the-size-of-the-economy-gross-domestic-product-cnx
7. Aliu A, Rexhepi S, Iseni E. Analysis and comparison of commitment, homework, extra hours, preliminary grades and testing of students in Mathematics using linear regression model. Mathematics Teaching Research Journal. 2021;13(3):21-52.
8. Macroeconomics: A european text (Michael C. Burda, Charles Wysplosz – published by Oxford University Press in; 2017. Available:https://shorturl.at/dKPRT

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/112671