LAUTECH Journal of Computing and Informatics (LAUJCI) – ISSN: 2714-4194
Volume 3 Issue 1, June 2023 – www.laujci.lautech.edu.ng

DEVELOPMENT OF A PREDICTIVE MODEL FOR PETROLEUM PRODUCTS DISTRIBUTION IN NIGERIA USING MACHINE LEARNING APPROACH

Famitimi R.F
Dept of Mathermatical & Computer Sciences University of Medical Sciences Ondo, Nigeria
rfamitimi@unimed.edu.ng

Emuoyibofarhe O.N
Computer Science Programme, Bowen University Iwo, Nigeria
ozichi.emuoyibofarhe@unimed.edu.ng

Adebayo S.
Mechatronics Programme, Bowen University Iwo, Nigeria
segun.adebayo@bowen.edu.ng

Ademoye A.A
Computer Science Programme, Bowen University Iwo, Nigeria
adetoye.ademoye@bowen.edu.ng

Awoniran T.I
Computer Science Programme, Bowen University Iwo, Nigeria
temitayo.awoniran@bowen.edu.ng

ABSTRACT

Petroleum products are the major economy of Nigeria today. The country being a producer and an exporter of these products is having internal challenge of availability of the product periodically due to inappropriate distribution coupled with the incessant increase in pump prices. This challenge has been on for decades. With this background information this research used a set of previous data set obtained from Nigeria National Petroleum Corporation and National Population Census data to develop a predictive model for the distribution of the product with the use of machine learning approach. Python programming tools were used for establishing possible relationship between the data sets before the development and implementation of the model. The developed model was evaluated and the error level was found to be tolerable this research provided a realistic model that can be used for the distribution of the product by the stake holders so as to address this age long challenge.

Keywords: Predictive Model, Machine Learning, Python Programming, Petroleum Products. Model Training

1. INTRODUCTION

Nigeria is said to have a huge deposit of petroleum products. It regained its position as Africa's largest oil producer in May, 2023 (Izuaka, 2023) by producing an average of 1.184 million barrels per day (Adeloju, 2023). Nigeria is the 15th world's largest exporter of crude oil as at May 2023. In spite of the fact that the country is among the producers and exporters of crude oil products, there has been an unending internal challenge of availability of the product which can be significantly traced to inappropriate distribution coupled with the incessant increase in pump prices (Anyaogu and Moyinoluwa, 2022; Olayungbo and Ojeyinka, 2022). As a result of this aforementioned challenge in Nigeria, subsequent sections are devoted to the conceptual overview, methodologies, methods, employed in addressing this challenge as well as the evaluation of the proposed result.

2. LITERATURE REVIEW

2.1 Machine Learing

Accordingr to Jordan and Mitchell (2016) , Machine learning is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science. Machine learning which is a subset of artifial intelligence helps humans to make decisions, based on data, with efficiency

and scale. Machine learning describes the capacity of systems to learn from problem-specific training data to automate the process of analytical model building and solve associated tasks Janiesch, Zschech and Heinrich (2021).

2.2 **Tools of Machine Learning**

In Machine learning, statistical tools and computer algorithms are used to make machine to learn from available or supplied data. The result of this is a model that can be applied to solve similar problems. A model accepts an input data in order be trained or when tested and gives an output. .In statistics, variables that can stand on their own are independent variables are hence referred to as input variables. On the other hand, dependent variables are referred to as outcome variables after being processed by the model. Also, independent (input) variables are also referred to as features and dependent variables referred to as target outcomes in machine learning (ML). When the outcome of a model is suggesting a future trend or an outcome from previously unknown data, it is generally termed a Predictive model (Manika and Madhusudhan, 2022; Padilla, 2019; Fjeld *et. al.* 2020). When Machine learning, statistical tools and computer algorithms are employed together with available data on Nigeria petroleum supply, sample population figures and the use of necessary utilities in Python, the distribution challenge will be greatly addressed.

2.3 **The Process of Machine Learning**

Machine learning which is an act of making the computer to learn from studying data and statistics is a step into the direction of artificial intelligence (AI). It can be conceived as a program that analyses data and learns to predict the outcome using some established methods. Before reliable prediction can be made based on existing data sets, there must be relationship, "r" between the two data sets. If there is no relationship also known as correlation, the data sets cannot be used to come up with a prediction for future data (Lina 2018; Mohammed, Khan and Bashier Mohammed 2016; Khadse V, Mahalle PN, Biraris SV 2018).

2.31 The relationship 'r' and interpretation

Python has inbuilt tools that can be used to ascertain whether a particular data set has relationship with another set. The "r" value for two data sets ranges from -1 to 1. If the value is 0, it means there is no relationship, if it is 1 this shows 100% positively related, and -1 means 100% negatively related. Python and the Scipy module have utilities for computing this value of 'r' after supplying the values of x and y.

2.32 Training and testing the model

If a relationship has been established, then the data sets will be divided into training and testing subsets in ratio 80% and 20% or 70% and 30% respectively. The training sets will then be used to obtain the proposed model depicted by Equation (1) while substituting Equations (2) and (3). Python utilities are available for evaluating Equation (1) with the use of Equations (2) and (3). When the model has been obtained, the remaining testing data will be used for testing.

$$Y = aX + b \quad \ldots\ldots \qquad \text{Equation (1)}$$

Where

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

$$b = \frac{[n(\sum xy) - (\sum x)(\sum y)]}{[n(\sum x^2) - (\sum x^2)]} \qquad \ldots \quad \text{Equation (3)}$$

3. METHODOLOGY

`

3.1 Data Sources

The data sources used for the study are the 2010 Annual Statistical Bulletin of Nigeria National Petroleum Corporation (NNPC) and 2010 National Population figures according to states. From these sources, data from ten (10) states were randomly selected for model training and testing. From the selection 70% of the data were used for training and the remaining 30% were used for testing. Table 1 contains the data selected from NNPC Monthly Petroleum distribution for the selected states, while Table 2 contains the National Population figures for the selected states.

Table 1. NPC Population Figures

Sno	States	Population
1	Kebbi	3,256,000
2	Kogi	3,314,000
3	Kwara	2,365,000
4	Lagos	9,114,000
5	Zamfara	3,278,000
6	Rivers	5,198,000
7	Ogun	3,751,000
8	Ondo	3,461,000
9	Osun	3,417,000
10	Oyo	5,581,000

Table 2. NNPC Monthly Petroleum distribution

Sno	States	Litters
1	Kebbi	45,169,000
2	Kogi	88,249,000
3	Kwara	48,773,000
4	Lagos	635,351,000
5	Zamfara	45,587,000
6	Rivers	170,023,000
7	Ogun	40,925,000
8	Ondo	77,616,000
9	Osun	100,764,000
10	Oyo	183,031,000

The script for checking the relationship is shown in figure 1, the one for extracting training and testing data is shown in figure 2. Figure 3 contains Python Script for Training of model, Figure 4 contains Python Script for Mathematical model extraction and Figure 5 contains Python Script for extracting evaluation data.

3.2 Checking the Relationship

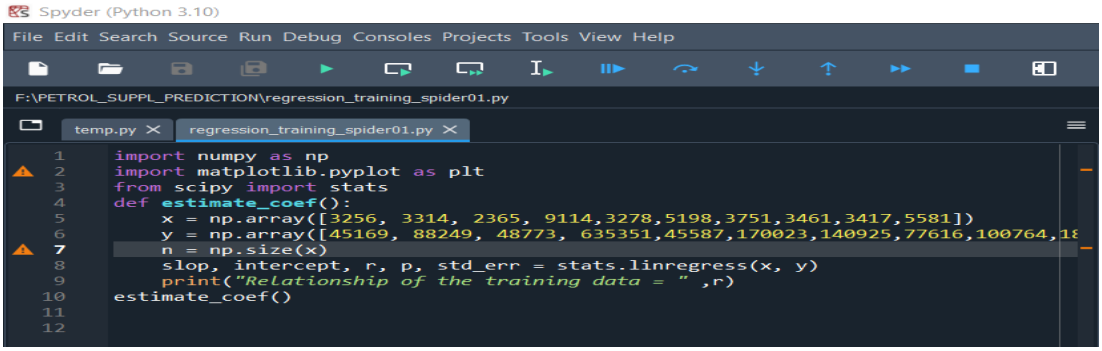

Figure 1. Python script for checking relationship of data sets.

3.3 Extracting Training Data Sets

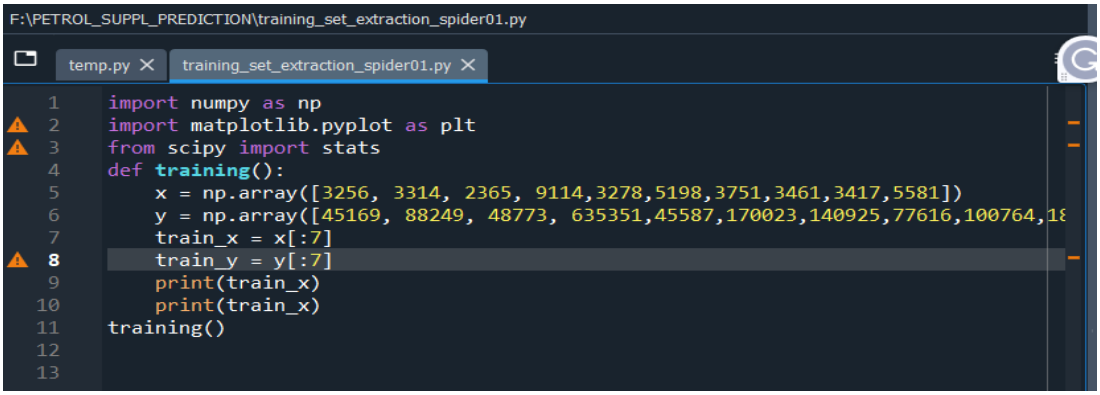

Figure 2. Python script for extracting the training data sets

3.4 Training of model

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
def model():
    x = np.array([3.25,3.31,2.36,9.11,3.27,5.19,5.58])
    y = np.array([45.17,88.24,48.77,635.35,45.59,170.02,183.031])
    n = np.size(x)
    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x
    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - ( b_1 * m_x )
    plt.scatter(x,y)
    # predicted response
    y_pred = b_0 +  (b_1 * x)
    slop, intercept, r, p, std_err = stats.linregress(x, y)
    print(r)
    plt.plot(x, y_pred)
    plt.xlabel('POPULATION')
    plt.ylabel('PETROLEUM CONSUPTION')
    plt.show()
model()
```

Figure 3. Python Script for Training of model

3.5 Script for Mathematical model extraction

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
def model_mathematical():
    x = np.array([3.25,3.31,2.36,9.11,3.27,5.19,5.58])
    y = np.array([45.17,88.24,48.77,635.35,45.59,170.02,183.031])
    n = np.size(x)
    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x
    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - ( b_1 * m_x )
    # predicted response
    y_pred = b_0 +  (b_1 * x)
    print("The Mathematical Model    = ")
    print(b_0, "  +  ",b_1,"x" )

model_mathematical()
```

Figure 4. Python Script for Mathematical model extraction

3.6 Extraction of evaluation data

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
def model_error():
    x = np.array([3.25,3.31,2.36,9.11,3.27,5.19,5.58])
    y = np.array([45.17,88.24,48.77,635.35,45.59,170.02,183.031])
    n = np.size(x)
    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x
    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - ( b_1 * m_x )
    # predicted response
    x1 = float(input("Input x the pupolation to be predicted "))
    y1 = float(input("Input y the expected result" ))
    y_pred = b_0 +  (b_1 * x1)

    print("If input is " , x1, )
    print("The Model will predict ", y_pred)
    print("Since the expected figure is ", y1)
    print("The percent error is ", ((y1 - y_pred)/y1)* 100)
model_error()
```

Figure 5. Python Script for extracting error data

4. RESULTS

4.1 Relationship checking result

The result obtained for relationship checking on the Population and Distribution data sets is shown in figure 6.

```
In [1]: runfile('F:/PETROL_SUPPL_PREDICTION/
regression_training_spider01.py', wdir='F:/
PETROL_SUPPL_PREDICTION')
Relationship of the training data =  0.9611420573265286
```

Figure 6. Relationship checking result

Since the relationship is within -1 and 1, it means the data set are related. Also because the value is close to 1, it shows a very good positive relationship. With this result, a regression analysis can be constructed.

4.2 Training model result

The result of the model training is shown in figure 7.

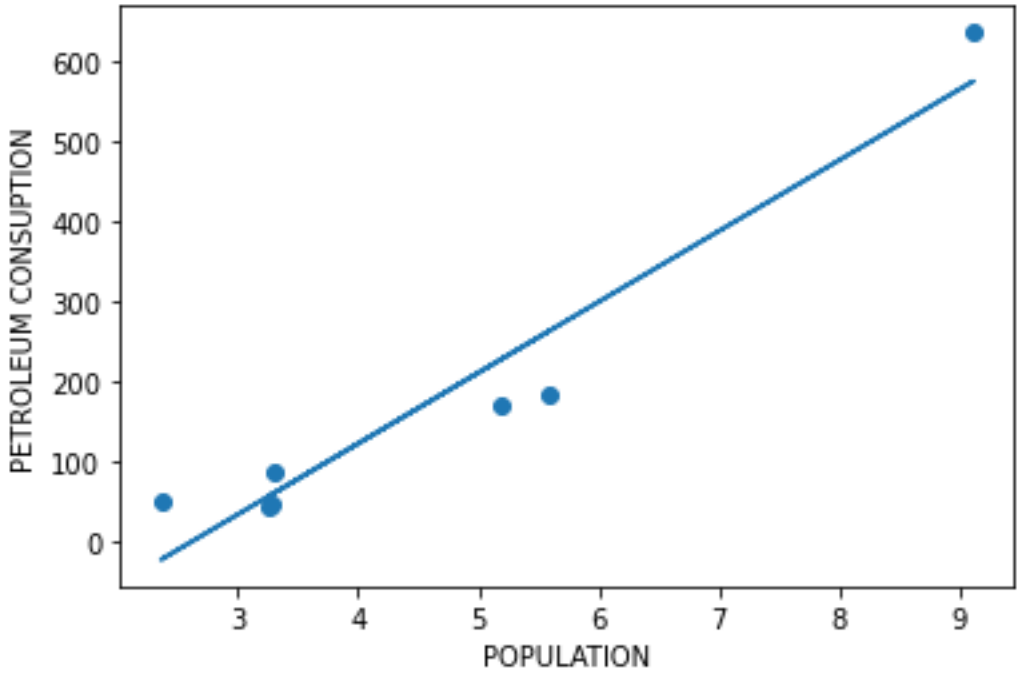

Figure 7. Model training result

4.3 Mathematical model extraction

The mathematical model is shown in figure 8.

```
In [35]: runfile('F:/PETROL_SUPPL_PREDICTION/
mathematical model extraction script spider01.py',
wdir='F:/PETROL_SUPPL_PREDICTION')
The Mathematical Model    =
-231.45237117237676    +    88.4420828876407 x
```

Figure 8. Mathematical model

4.4 Model testing

The mathematical model testing using the testing data is shown in figure 9

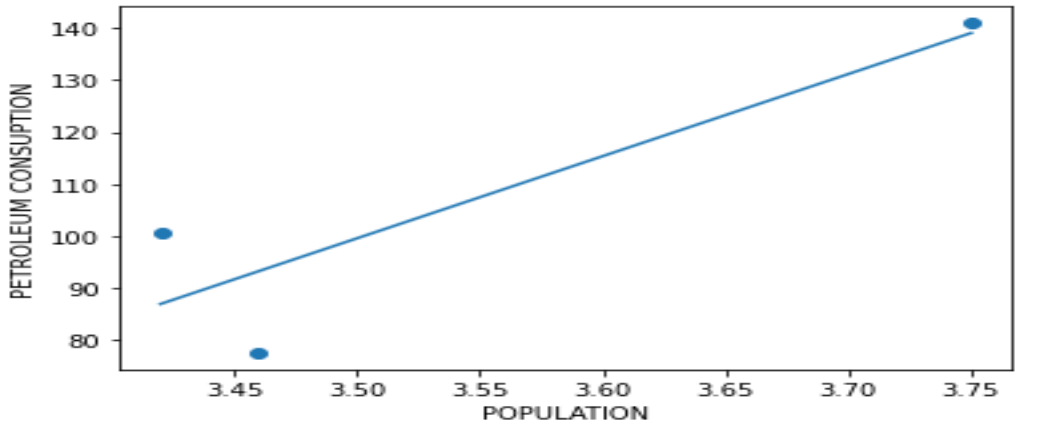

Figure 9. Model testing using testing data

4.5 Model evaluation

The model is evaluated by computing percentage errors on the testing data (figures 10) and then shown table 3 for the purpose of Mean Absolute Error (MAE) computation.

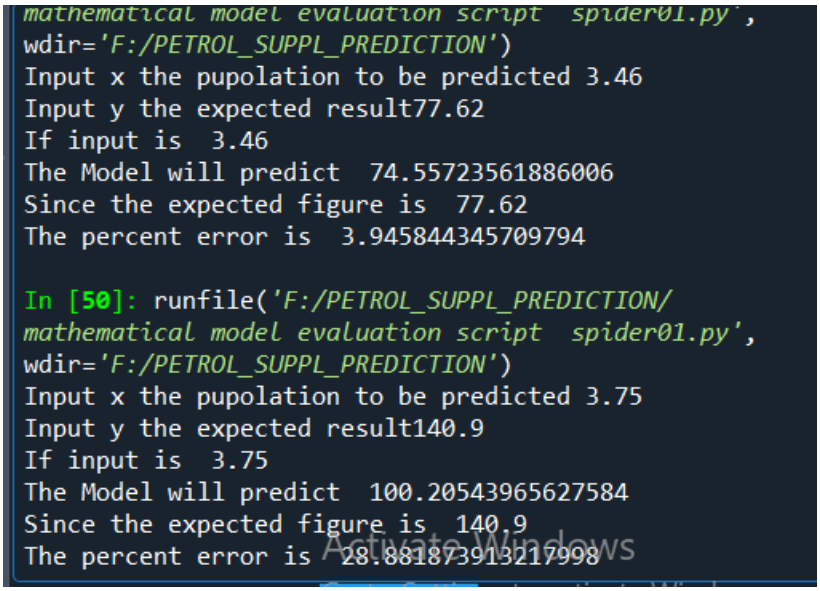

Figure 10. Model Error Computation

Table 3: Error Computation Table

Population (Million)	**Expected Distribution (Million)**	**Predicted Distribution (Million)**	**Error Percent**
3.46	77.62	74.56	3.94%
3.75	140.9	100.20	28.88%
3.42	100.76	71.02	29.52%
Mean Absolute Error (MAE)			20.78%

5. CONCLUSION

The developed model has shown a very good relationship between state population data and distribution data by showing a value of 0.96. With this value, a regression can be used for prediction of the expected petrol distribution to any state in Nigeria. Also error level of 20.78% is tolerable and it shows that the accuracy of the model is 79.22%. This research provided a realistic model that can be used for the distribution of the product by the stake holders so as to address this age long challenge of distribution irregularities

6. ACKNOWLEDMENTS

Our thanks to the management of Bowen University Iwo and University of Medical Sciences Ondo that provided us support towards success of this work.

7. REFERENCES

[1] Adeloju B. (2023). Nigeria regains spot as. Africa's largest oil producer. The Cable. News and views unlimited. June 15, 2023.

[2] Anyaogu I. and Moyinoluwa A (2022). In a speech at the 2022 edition of OTL Africa Week, in Lagos, themed: "Regulating downstream energy transition in dynamic times"..

[3] Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. SSRN ElectronJ. https://doi.org/10.2139/ssrn.3518482

[4] Izuaka M. (2023). Nigeria regains position as Africa's largest oil producer. Premium Times. June 16, 2023.

[5] Janiesch C., Zschech P. and Heinrich K. (2021). Machine learning and deep learning. Springer. Nature.

[6] Jordan M.I. and Mitchell T. M. (2015). Machine learning: Trends, perspectives, and prospects SCIENCE sciencemag.org, VOL 349 ISSUE 6245 page 255

[7] Khadse V, Mahalle PN, Biraris SV. An empirical comparison of supervised machine learning algorithms for internet of things data. In: 2018 Fourth International Conference on Computing Communication Control and utomation (ICCUBEA), IEEE. 2018; 1–6

[8] Lina Klass (2018). Machine Learning – Definitions and applications Examples. https://www.spotlightmetal.com/Predictive maintainance-based-on-ai-a-891610/

[9] Manika Lamba and Madhusudhan Margam (2022). Predictive Modeling. In book: Text Mining for Information Professionals: An Uncharted Territory (pp.213-242).

[10] Mohammed M, Khan MB, Bashier Mohammed BE (2016). Machine learning: algorithms and applications. CRC Press; 2016

[11] Padilla T (2019) Responsible operations: Data science, machine learning, and AI in libraries. OCLC RESEARCH POSITION PAPER, Dublin, Ohio: OCLC Research. https://doi.org/10.25333/xk7z-9g97. Accessed 08 July 2023

Author's Brief Profile

Famutimi R. F. (PhD) is presently an Associate Professor of Computer Science in the Department of Mathematical and Computer Sciences, University of Medical Sciences, Ondo, Ondo State. He holds B.Sc, M.Sc, M.Phil and PhD all in Computer Science. His research areas include Data Science, Machine Learning and Software Engineering. He can be reached by phone on +2348035894479 and through E-mail rfamutimi@unimed.edu.ng.