OPTIMIZATION AND PREDICTION OF CROP (MAIZE) YIELD WITH MACHINE LEARNING IN SUB-SAHARAN AFRICA (NORTH-EASTERN, NIGERIA)

Ph.D. Proposal

Presented by Ezra Daniel Dzarma (Mat. 40503623)

@

General Seminar of Institut de Mathématques et de Sciences Physiques (IMSP)

Filiére: Technologies de l'Information et de la Communication (TIC)

Supervisors:

Prof. Degla Guy & Prof. Theophile Komlan Dagba

08th November, 2023

Abstract

This Proposed Research is focused towards:

i. Applying CRISP DM
ii. Developing farm inputs optimization model
iii. Accurate prediction of crop yields

Bibliometric literature review was conducted on the area of study focusing on:

i. Authors
ii. Productivity
iii. Citations
iv. Collaboration

1. Background of the study

In food production, every nation relies on some common cereals. Corn is significantly useful in the production of :

i. feed,

ii. industrial raw materials,

iii. and bioenergy

In some developing nations, maize is a staple meal that accounts for 30 to 50 of daily caloric consumption. Maize is used in highly processed goods

2. Background of the study Continues

Crop cultivation in Africa faces several challenges that can hinder agricultural productivity and food security. Some of the key problems include (Sanou & Andersson 2019):

i. climate change

ii. Policy and institutional challenges

iii. Limited access to irrigation

iv. Soil degradation

v. Limited access to finance and resources

vi. Post-harvest losses

3. Background of the study continues

- Several researchers have made efforts to model maize production and yield, few among them are:

  i. Fashoto *et al.* (2021)

  ii. Ornella *et al.* (2012)

  iii. AONO *et al.* (2022)

  iv. Bi *et al.* (2022),

- Despite all these tremendous effort maize production in Africa has not met the demand of consumers

4. Statement of a Problem

i. Nigeria is the most populated nation in Africa and one of the of the nations that have invested heavily in maize production.

ii. Despite this, the nation still has difficulties producing enough food Sanou and Andersson (2019).

iii. Many efforts had been made to address the challenges of food security to no avail (Boussard *et al.*, 2006, Hilderink *et al., 2012*)

iv. Hence, this research would be focusing towards developing a mathematical models and ML algorithm that would optimize maize production and accurately forecast crop yields

5. Aim and Objectives

The aim of this proposed research is to:
develop farm inputs optimization model and determine accurate ML prediction algorithm

The objectives are to:

i. to apply CRISP DM

ii. Modify Cobb-Douglas production function

iii. minimize the cost of production;

iv. maximize farm output;

v. Forecast crop yields accurately; vi. Provide information to farmers

6. Justification of the Study

i. Food Security

ii. Flexible framework

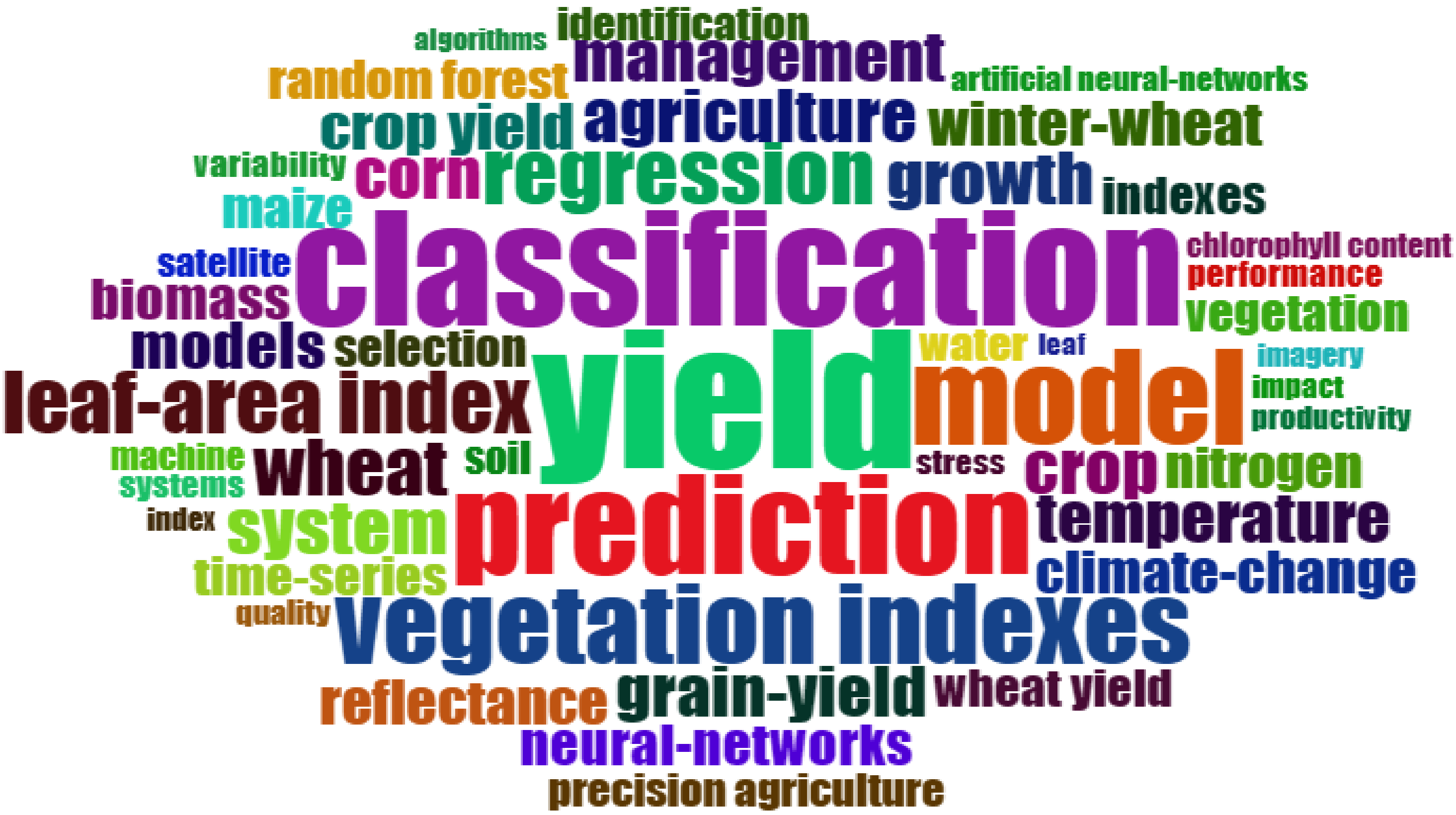
iii. Good fit with real-world data

iv. Adaptability to changing conditions

7. BIBLIOMETRIC LITERATURE REVIEW OF OPTIMIZATION AND PREDICTION OF CROP YIELDS

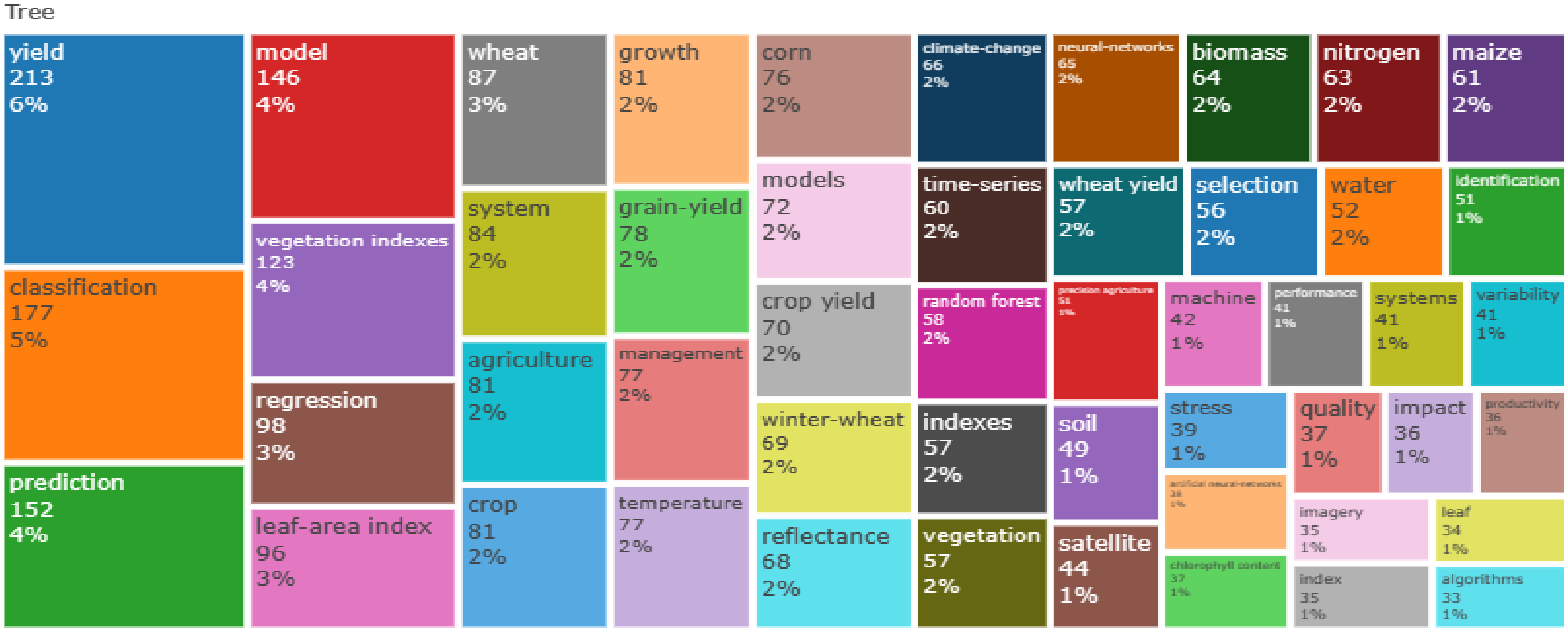
i. Bibliometric data were extracted from the Web of Science.
ii. Software called Bibliometrix applied to analyse the data
iii. Time frame considered 2012-12/05/2023
iv. Annual growth rate 47.14%
v. Authors: 5115
vi. 12 single-authored writers
vii. 12 single-authored documents
viii. 6 co-authors per document.

8. Word Cloud

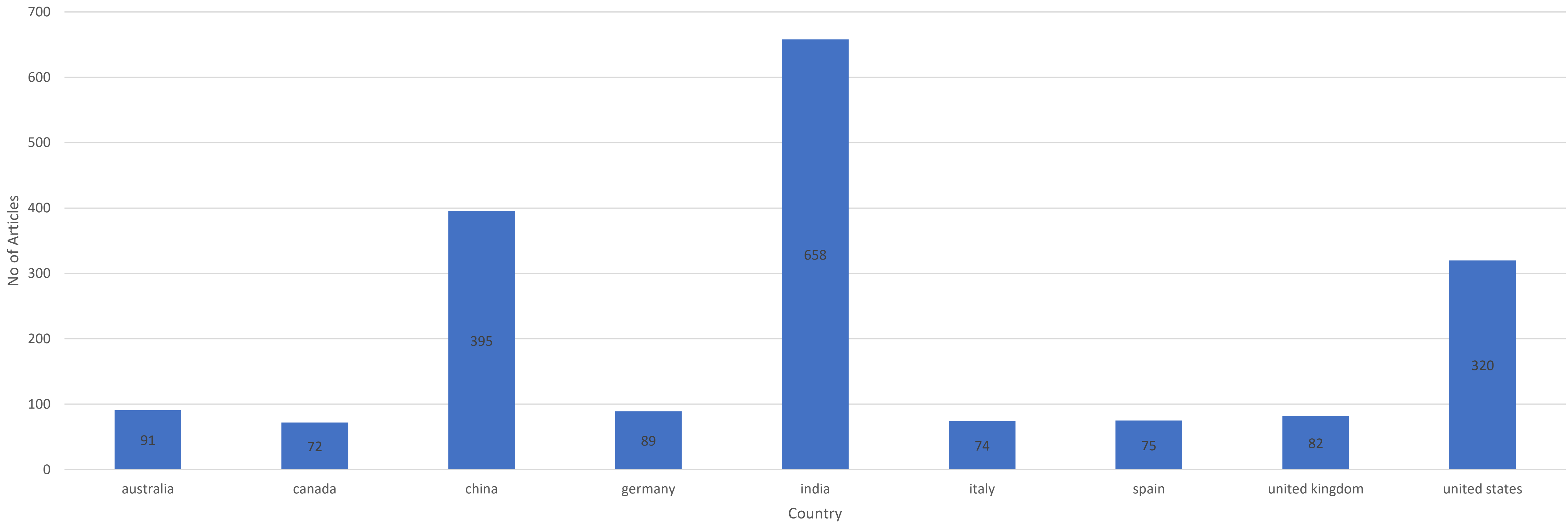

algorithms
identification
management
random forest
artificial neural-networks
crop yield
agriculture
winter-wheat
variability
corn
regression
growth
indexes
maize
classification
chloropyll content
performance
satellite
biomass
vegetation
models
selection
water
leaf
imagery
leaf-area index
yield
model
impact
productivity
machine
systems
wheat
soil
stress
crop
nitrogen
index
system
prediction
temperature
time-series
climate-change
quality
vegetation indexes
reflectance
grain-yield
wheat yield
neural-networks
precision agriculture

-

9. Word Tree


Tree
yield 213 6%
classification 177 5%
prediction 152 4%
model 146 4%
vegetation indexes 123 4%
regression 98 3%
leaf-area index 96 3%
wheat 87 3%
system 84 2%
agriculture 81 2%
crop 81 2%
growth 81 2%
grain-yield 78 2%
management 77 2%
temperature 77 2%
corn 76 2%
models 72 2%
crop yield 70 2%
winter-wheat 69 2%
reflectance 68 2%
climate-change 66 2%
time-series 60 2%
random forest 58 2%
indexes 57 2%
vegetation 57 2%
neural-networks 65 2%
wheat yield 57 2%
precision agriculture 51 1%
soil 49 1%
satellite 44 1%
biomass 64 2%
selection 56 2%
nitrogen 63 2%
water 52 2%
maize 61 2%
identification 51 1%
machine 42 1%
performance 41 1%
systems 41 1%
variability 41 1%
stress 39 1%
quality 37 1%
impact 36 1%
productivity 36 1%
artificial neural-networks 38 1%
imagery 35 1%
leaf 34 1%
chlorophyll content 37 1%
index 35 1%
algorithms 33 1%

10. Country's Productivity



Benin Republic 1 document and Nigeria 18 documents

11. RESEARCH METHODOLOGY

- The methodology for this study shall involve at least four approaches (Banua & Geetha, 2021).

  i. CRISP DM

  ii. Modified Cobb-Douglas production function

  iii. Genetic Algorithm

  iv. Random Forest

  v. K-Nearest Neighbour

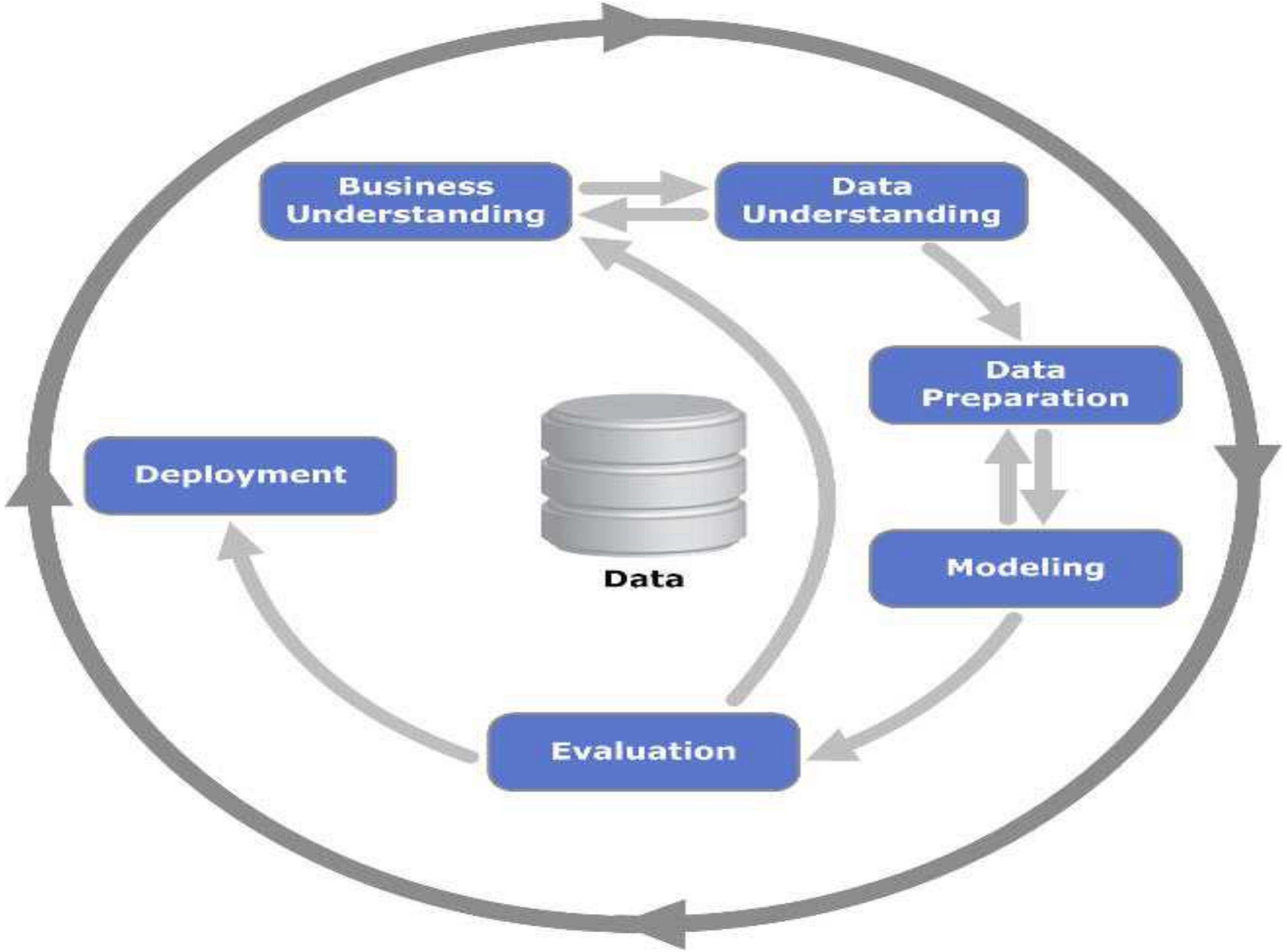
  vi. Neural Network

i. BG-Riceyield randomforestneuralnetwork

Data Collection

- The research would make use of two categories of data.
- The primary data (Hox & Boeije, 2005)
- The secondary data (Boslaug, 2007)

13. The CRISP DM Methodology



Source: Chapman (2000)

14. Optimisation Model

- Cobb-Douglas Production

- $Y = AK^{\beta}L^{\beta-1}$ Equ. 1

- Y equals total output,
- K is capital input
- L is labour input
- A is TFP

Cottrell (2019)

L
D

15. Cobb-Douglas Production Continues

Cobb-Douglas production function with multiplicative error term is formulated by Nurunnajib, *el al., 2018 as:*

$$Q_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} e_t$$

Equ. 2

Cobb-Douglas function with additive error term is formulated with:

$$Q_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} + e_t$$

Equ. 3

The output elasticity of capital (E_K) is measured through:

$$E_K = \frac{\%\,\Delta Q}{\%\,\Delta K} = \beta_3$$

Equ. 4

16 Cobb-Douglas Production Continues

While the output elasticity of labor (E_L) can be measured through:

$$E_L = \frac{\%\,\Delta Q}{\%\,\Delta L} = \beta_2$$

Equ. 5

Return to Scale

Suppose K_0 and L_0 that use of and will produce output , that is:

$$Q_0 = \beta_1 L_0^{\beta_2} K_0^{\beta_3}$$

Equ. 6

17. Cobb-Douglas Production Continues

- If Q_1 is an output generated by a combination of capital inputs, then it is obtained:

$$Q_1 = \beta_1 L_0^{\beta_2} (2K_0)^{\beta_3} = 2^{\beta_3} Q_0$$

Equ. 7

If Q_2 is the output produced by a combination of labor input, then it is obtained:

$$Q_2 = \beta_1 (2L_0)^{\beta_2} K_0^{\beta_3} = 2^{\beta_2} Q_0$$

Equ. 8

18. Cobb-Douglas Production Continues

- If Q_3 is the output produced by a combination of capital and labor inputs, then it is obtained:

$$Q_3 = \beta_1(2L_0)^{\beta_2}(2K_0)^{\beta_3} = 2^{\beta_2+\beta_3}Q_0$$

Equ. 9

1. If $\beta_2 + \beta_3 = 1$ then the function will show constant return to scale.

2. If $\beta_2 + \beta_3 < 1$ then the function shows the decreasing returns to scale

3. If $\beta_2 + \beta_3 > 1$ then it shows the scale with increasing return to scale

19. LIMITATIONS OF COBB-DOUGLAS PRODUCTION FUNCTION

i. It does not have constraint equation

ii. It assumes constant return to scale

iii. Assumption of constant technology: technology is a dynamic concept

iv. All the inputs include in the function must be positive. If one of the factor inputs of the function are zero, total output will be zero.

20. Modified Cobb Douglas Optimization Model

$Maximize(R(b, f, h, l, p, s) = Ab^{\propto_1} f^{\propto_2} h^{\propto_3} l^{\propto_4} p^{\propto_5} s^{\propto_6}$ (Objective function)) eqn. 10

Constraint Equations

$k_2 f + k_3 h + k_5 p \leq q_2$ (Chemical con.) eqn. 11

$k_1 b \leq q_3$ (labour con.) eqn.12

$k_6 s \leq q_3$ (Seed con.) eqn.13

l, h, f, b, s ≥ 0

$where \propto_1, \propto_2, \propto_3, \propto_4, \propto_5, , \propto_6$ are positive parameters while A, b, f, h, l, p, and s are technological factors, land, herbicides, fertilizer, labour and Seed respectively.

Maize Input Model Optimization for Yobe State

$Maximize(R(b, f, h, l, p, s) = 2.259 * l^{0.2022} b^{1.903} s^{0.727} h^{0.3622} p^{0.654} f^{0.7129}$ (Objective function)) eqn. 14

Constraint Equations

$4.3f + 6h + 3.5p \leq q_2$ (Chemical con.) eqn. 15

$b \leq 30$ (labour con.) eqn.16

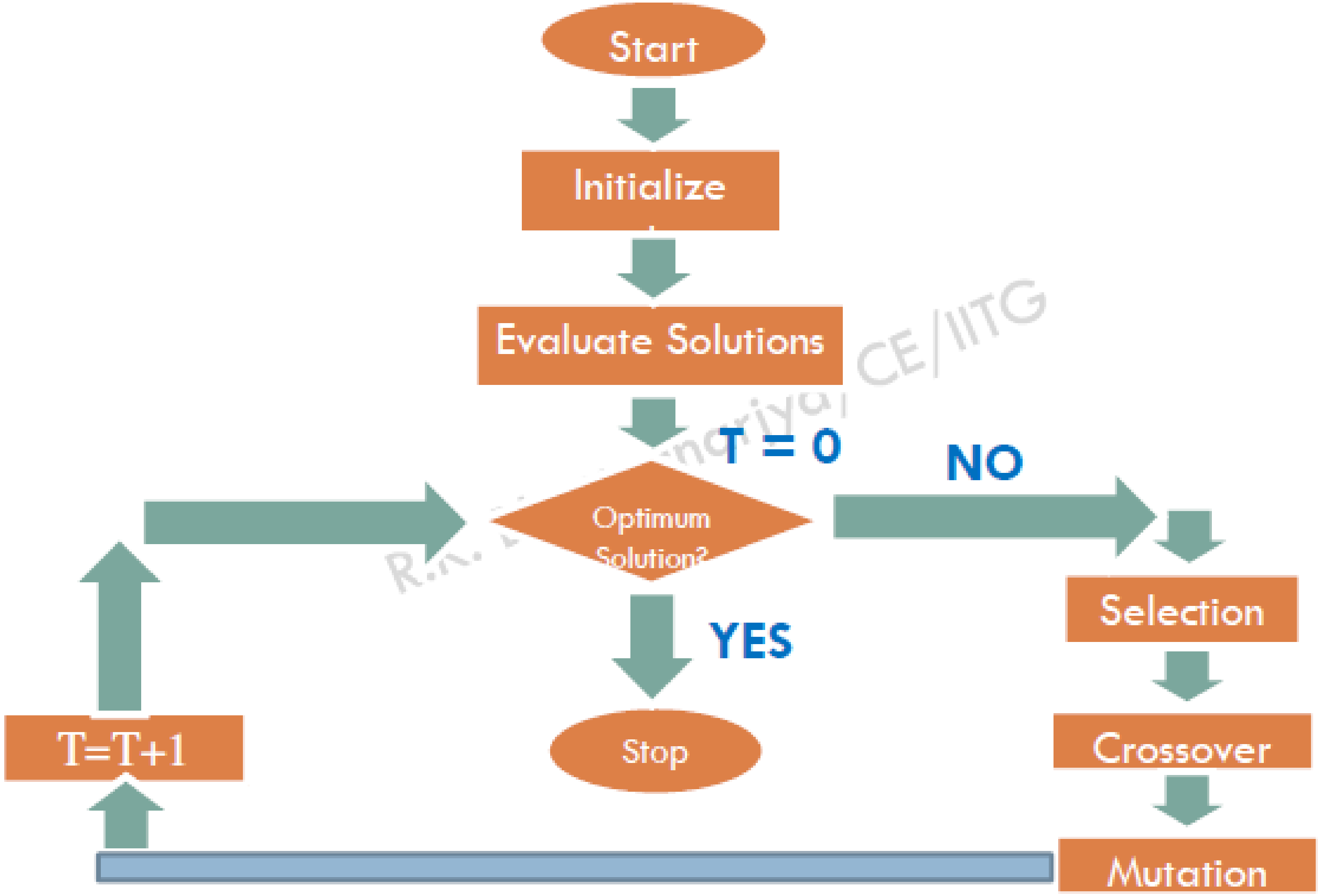
$s \leq 38.57$ (Seed con.) eqn. 17

$l, h, f, b, s \geq 0$

$where \propto_1, \propto_2, \propto_3, \propto_4, \propto_5, , \propto_6$ are positive parameters while A, b, f, h, l, p, and s are technological factors, land,

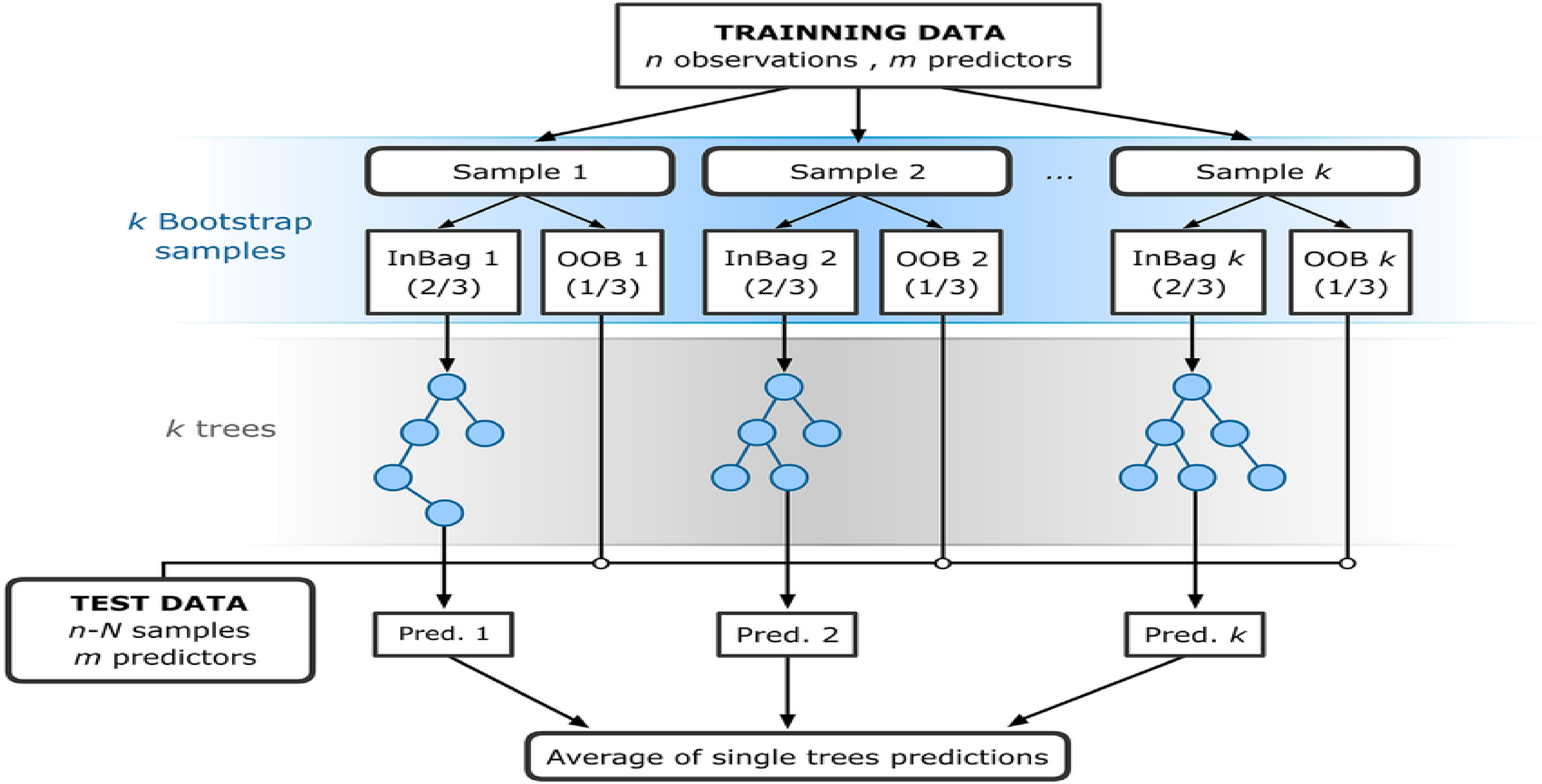
herbicides, fertilizer, labour and Seed respectively.

21. Simple Genetic Algorithm



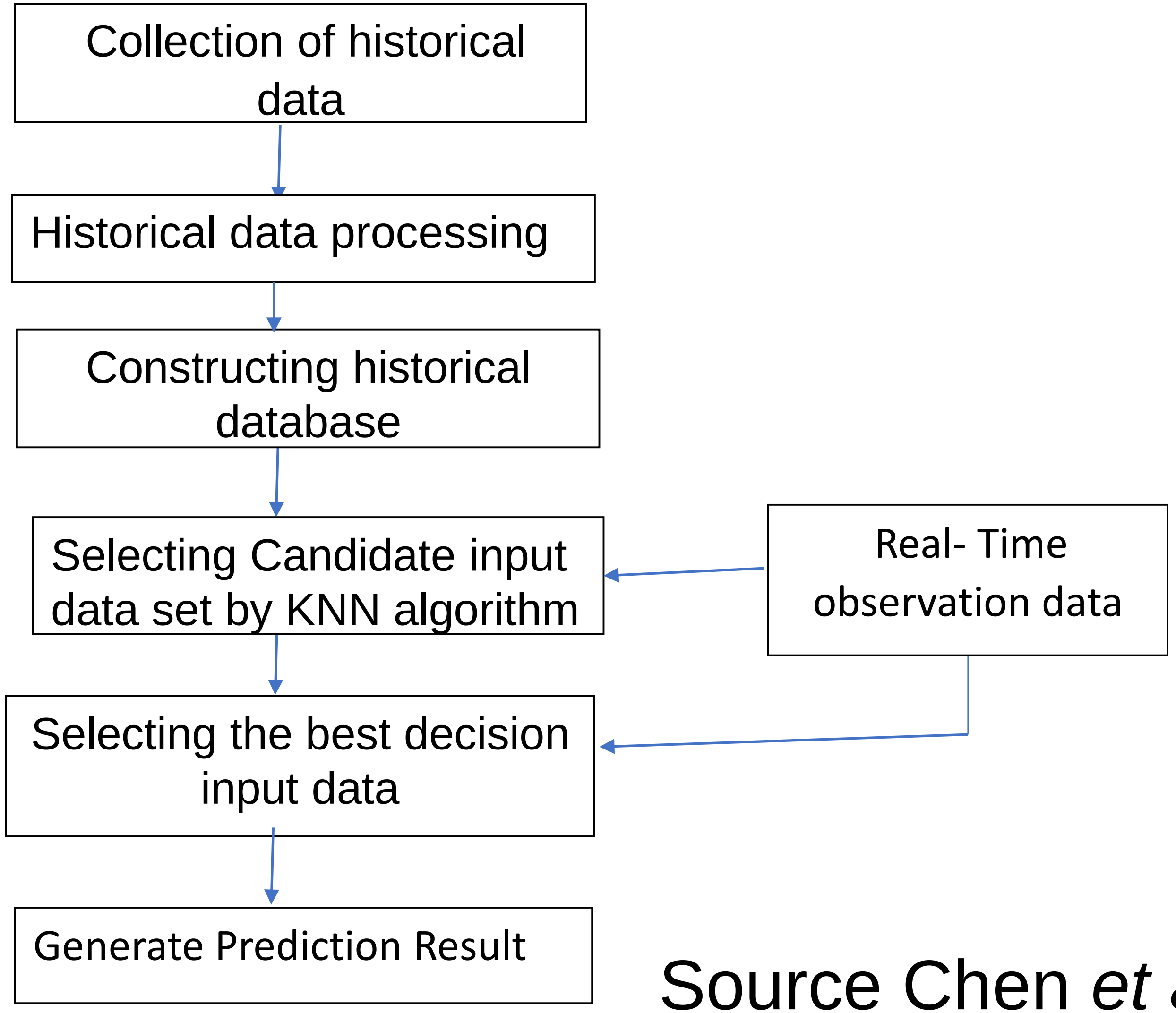
Source: Bhattacharjya, R. K. (2012)

22. Random Forest Method



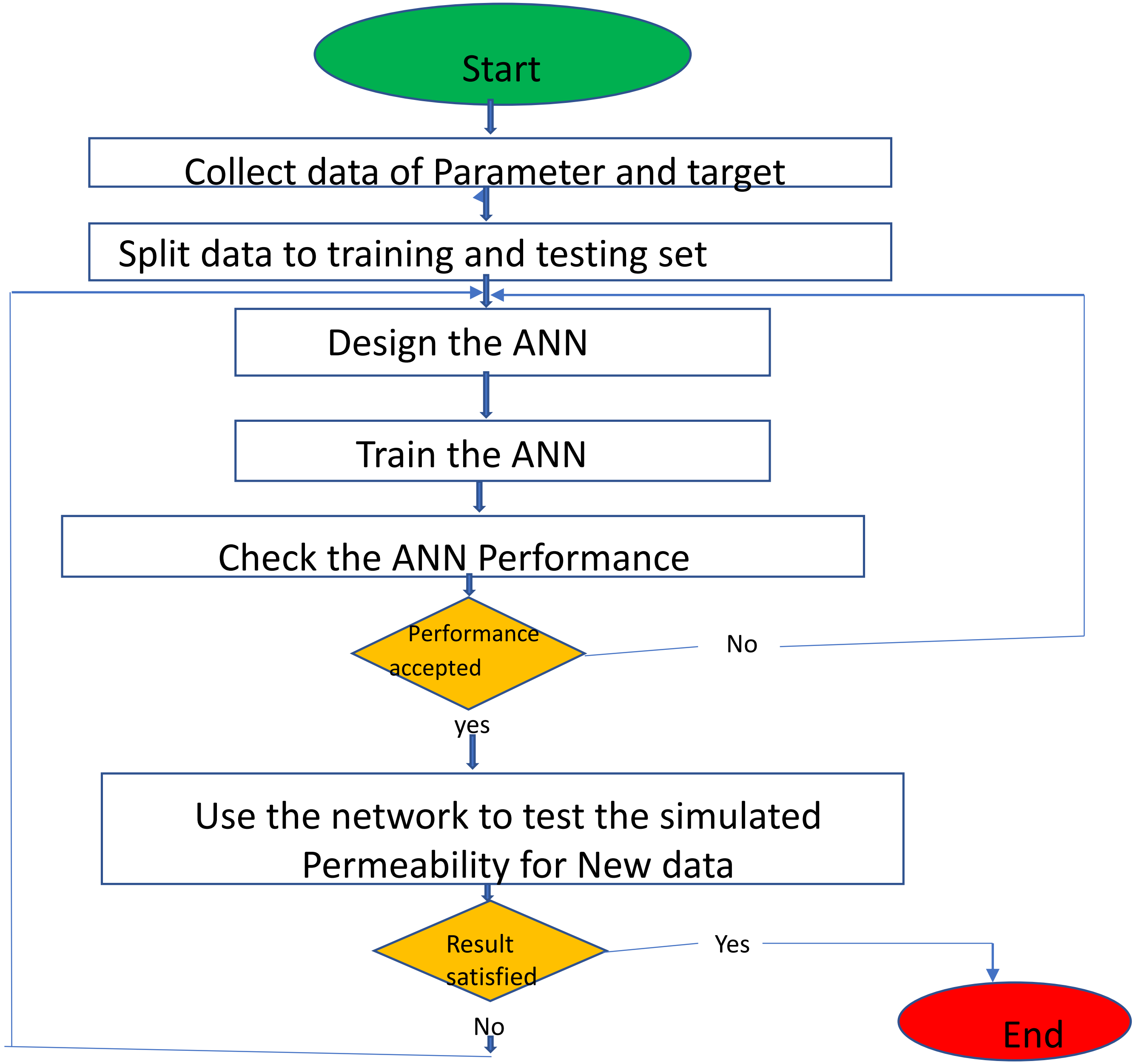
Source: Rodriguez-Galiano *et al* (2016)

23. Flow chart of KNN nonparametric regression algorithm



Source Chen *et al* (2019)

24. NNs methodology flow chart



Source: Mahdi & Holdick (2017)

25. REFERENCES

1. Aono, A. H., Pimenta, R. J. G., Francisco, F. R., De souza, A. P., & Lorena, A. C. (2022). Machine learning for crop science: applications and perspectives in maize breeding. *revista brasileira de milho e sorgo*, *21*.

2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Banua, E., & Geetha (2021). A. Rice crop yield prediction using random forest and deep neural network-An integrated approach.

3. Bi, C., Hu, N., Zou, Y., Zhang, S., Xu, S., Yu, H., 2022. Development of deep learning methodology for maize seed variety recognition based on improved skin transformer. Agronomy 12, 1843.

4. Boslaugh, S. (2007). An introduction to secondary data analysis. *Secondary data sources for public health: A practical guide*, 2-10.

5. Boussard, J. M., Daviron, B., Gérard, F., & Voituriez, T. (2006). Food Security and Agricultural Development in Sub-Saharan Africa. *FAO: Rome, Italy*.

26. NEXT LINE ACTION

- 1. Complete Primary data collection in North-East Nigeria
- 2. Obtain data of climatic factors that influences Crop yield across Africa

3. Develop farm inputs optimization model to each state of the North-East Nigeria

4. Apply machine learning techniques

27. Challenges

- 1. Time constraint

- 2. Financial Constraints

2. Security Challenges

6. Cottrell, A., (2019). The Cobb-Douglas Production Function. Economics 207, 1–4.

7. Chen, K., Zhao, S., & Zhang, D. (2019, June). Short-term Traffic Flow Prediction based on Data-Driven Knearest neighbour Nonparametric Regression. In *Journal of Physics: Conference Series* (Vol. 1213, No. 5, p. 052070). IOP Publishing.

8. Bhattacharjya, R. K. (2012). Introduction to genetic algorithms. *IIT Guwahati*, *12*.

9. . Fashoto, S.G., Mbunge, E., Ogunleye, G., den Burg, J.V., 2021. Implementation of machine learning for predicting maize crop yields using multiple linear regression and backward elimination. Malays. J. Comput. MJoC 6, 679–697.

REFERENCES

10. Hox, J. J., & Boeije, H. R. (2005). Data collection, primary versus secondary.

11. Hilderink, H. B. M., Brons, J., Ordonez, J., Akinyoade, A., Leliveld, A. H. M., Lucas, P., & Kok, M. T. J. (2012). Food security in sub-Saharan Africa: An explorative study.

12. Kralovec, S. (2020). Food insecurity in Nigeria-An analysis of the impact of climate change, economic development, and conflict on food security.

13. Mahdi, F.M. & Holdich. R. G. (2017) Using statistical and artificial neural networks to predict the permeability of loosely packed granular materials. Separation Science and Technology, 52:1, 1-12, DOI: 10.1080/01496395.2016.1232735

Nurunnajib, A. F., Wulan, E. R., Awalluddin, A. S., Supian, S., & Subiyanto, S. (2018). Application of Cobb-Douglas production function to manufacturing industries in West Sumatra Indonesia. *World Scientific News*, (101), 145-156.

28. REFERENCES

14. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS inc*, *9*(13), 1-73.

15. Ornella, L., Cervigni, G., and Tapia, E. (2012). Applications of machine learning for maize breeding. French Argentine International Center for Information and Systems Sciences, 27 de Febrero 210

16. Rodriguez-Galiano, V. F., Sanchez-Castillo, M., Dash, J., Atkinson, P. M., & Ojeda-Zujar, J. (2016). Modelling interannual variation in the spring and autumn land surface phenology of the European forest. *Biogeosciences*, *13*(11), 3305-3317.

17. Sanou, J., & Andersson, L. (2019). Enhancing smallholder farmers' access to finance in Sub-Saharan Africa: A review of recent trends, donor interventions, and impact. World Development, 122, 610-622.

18. West Africa Gateway (WAG)

Source: www.westafricagateway.or, retrived 03/07/2023

Thank You for Listening

Regional Coordination Unit (RCU)

International Centre of Insect Physiology and Ecology (icipe)

P.O. Box 30772 – 00100, Nairobi, Kenya

Tel +254 (20) 8632000

Email: rsif@icipe.org

Website: www.rsif-paset.org