# Abstract

Sustainable agriculture plays a pivotal role in achieving food security, addressing hunger, and promoting responsible resource management, as outlined in Sustainable Development Goal 2. This report examines the indicators of agricultural factor income per AWU (Annual Work Unit), **Area under organic farming (%), and** Government support to agricultural research and development (R&D)as a measure of sustainable development in the European context. By leveraging machine learning techniques, including decision trees, k-nearest neighbors, and neural networks, we aim to uncover patterns and relationships within the dataset obtained from Pordata, which provides information on agricultural factor income across various European countries. The results shed light on the key variables that significantly impact sustainable agriculture, providing valuable insights for policymakers and stakeholders to make informed decisions towards achieving the related Sustainable Development Goals.

**Table of Contents**

Abstract……………………………………………...2

Introduction ………………………………………5-6

Methods……………………………………..……7-12

Results & Discussion……………………………13-25

Conclusion……………………………………….…26

References…………………………………………..27

Appendix………………………………………28-33

# 1.0 Introduction

# Achieving sustainable agriculture is crucial for ensuring food security, addressing malnutrition, and promoting responsible resource management. This report investigates the indicators of agricultural factor income per AWU,**Area under organic farming, and** Government support to agricultural research and development (R&D) as a measure of sustainable development in the European context. By analyzing the dataset obtained from Pordata, which provides information on agricultural factor income across various European countries, we aim to identify the factors that significantly impact this indicator. Understanding these factors is essential for policymakers and stakeholders to make informed decisions toward promoting sustainable agriculture and achieving the related Sustainable Development Goals.

# 

# 1.1 Data Description:

# The dataset used in this analysis covers Europe, including Austria, Belgium, Bulgaria, Cyprus, Croatia, Denmark, Slovakia, Slovenia, Spain, Estonia, Finland, France, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Czech Republic, Romania, Sweden, Iceland, Norway, United Kingdom, and Switzerland. It provides information on agricultural factor income per AWU, measured as an index number with the base year of 2010. The utilization of the Annual Work Unit (AWU) concept enables standardized measurement of labor input across countries. In order to have more conclusions, I included two more indicators which are:

# ****Area under organic farming (%):** The indicator measures the share of total utilised agricultural area (UAA) occupied by organic farming (existing organically-farmed areas and areas in process of conversion).**

# **Area under organic farming during the calendar year / Total utilised agricultural area during the calendar year) \* 100**

## **Government support to agricultural research and development (R&D):** The indicator refers to data measure government support to research and development (R&D) activities, or, in other words, how much priority governments place on the public funding of R&D.

# 2.0 Methods

# To identify the key variables impacting agricultural factor income per AWU, I employ machine learning techniques, including decision trees, k-nearest neighbors, and neural networks. These methods allow for uncovering patterns and relationships within the dataset, facilitating an understanding of the factors influencing sustainable agriculture. By leveraging the power of machine learning, I can provide valuable insights for policymakers and stakeholders seeking to make informed decisions in support of sustainable agriculture and the related Sustainable Development Goals.

**2.1 Goal 2 Targets:**

**Annual Work Unit (AWU)-2.2** By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round.

**Government support to agricultural research and development (R&D)-2.3** By 2030, double the agricultural productivity and incomes of small-scale food producers, in particular women, indigenous peoples, family farmers, pastoralists and fishers, including through secure and equal access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment.

**Area under organic farming-2.4 (AUOF):**  By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters and that progressively improve land and soil quality.

**2.2 CODE:**

This is the step by step algorithm of the R-studio code used:

**Step 1: Load the Libraries**

library(ggplot2) # for graphics  
library(dplyr) # For piping function

library(rpart) # For decision trees  
library(class) # For k-NN  
library(neuralnet) # For neural networks

#install.packages('caTools')  
#install.packages('party')  
#install.packages('magrittr')  
#install.packages('rpart.plot')  
library(caTools)

library(magrittr)  
library(rpart.plot)

library(caret)

#install.packages(c('neuralnet','keras','tensorflow'),dependencies = T)

**Step 2: Load The Dataset**

df <- read.csv("AWU.csv",sep = "," ,header = TRUE)  
df <- as.data.frame(df)  
head(df)

**Step 3: Clean data by removing all empty cell**

# remove all rows with (x Not available )  
df <- subset(df, !apply(df == "x", 1, any))  
  
df["Target"] <- as.factor(x=df$Target)  
str(df) # check the structure of the cleaned data

sum(is.na(df))

# Step 4 : Statistical Summaries

summary(df)

# Step 5: Graphs For Few Selected countries

# Visualize the trend of indicators over time for Austria

ggplot(df, aes(x=Years , y = Austria, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Austria Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Belgium

ggplot(df, aes(x=Years , y =Belgium, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Belgium Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for United Kingdom

ggplot(df, aes(x=Years , y=UnitedKingdom, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For United Kingdom Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Spain

ggplot(df, aes(x=Years , y =Spain, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Spain Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Hungary

ggplot(df, aes(x=Years , y =Hungary, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Hungary Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Finland

ggplot(df, aes(x=Years , y =Finland, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Finland Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for France

ggplot(df, aes(x=Years , y =France, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For France Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Portugal

ggplot(df, aes(x=Years , y =Portugal, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For Portugal Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for CzechRepublic

ggplot(df, aes(x=Years , y =CzechRepublic, color = Target)) + geom\_line() +  
 labs(title = "Sustainable Development Goal 2 For CzechRepublic Over The Years", x = "Years", y = "Selected Indicators")

Visualize the trend of indicators over time for Poland

ggplot(df, aes(x=Years , y =Poland, color = Target)) + geom\_line()+  
 labs(title = "Sustainable Development Goal 2 For Poland Over The Years", x = "Years", y = "Selected Indicators")

**Step 6: Models Training**

set.seed(123) # For reproducibility  
train\_indices <- sample(1:nrow(df))  
df <- df[train\_indices,2:30] # shuffle the dataset excluding the years  
  
# Split data into training and testing sets  
create\_train\_test <- function(data, size = 0.8, train = TRUE) {  
 n\_row = nrow(data)  
 total\_row = size \* n\_row  
 train\_sample <- 1: total\_row  
 if (train == TRUE) {  
 return (data[train\_sample, ])  
 } else {  
 return (data[-train\_sample, ])  
 }  
}  
  
# feature scaling   
train\_data <- create\_train\_test(df, 0.8, train = TRUE)  
test\_data <- create\_train\_test(df, 0.8, train = FALSE)  
  
train\_data[,2:29] <- as.data.frame(scale(train\_data[,2:29]))  
test\_data[,2:29] <- as.data.frame(scale(test\_data[,2:29]))  
  
  
# Check the dimension of the data   
dim(train\_data)

dim(test\_data)

prop.table(table(train\_data$Target))

# Check for possible NA's   
sum(is.na(train\_data))

sum(is.na(test\_data))

head(train\_data)

# Decision Trees:

#### 

# Decision Tree  
tree\_model <- rpart(Target~., data = train\_data , control = rpart.control(maxdepth = 30,minsplit = 1))  
rpart.plot(tree\_model)

summary(tree\_model)

tree\_model$variable.importance

# Performance Evaluation  
tree\_predictions <- predict(tree\_model, test\_data, type = "class")  
tree\_accuracy <- (sum(tree\_predictions == test\_data$Target) / nrow(test\_data)) \* 100  
# Print the metrics  
cat("Decision Tree Accuracy:", tree\_accuracy, "% \n")

**K nearest neighbors:**

# Choose the K-Value for model accuracy.  
  
trainControl <- trainControl(method="repeatedcv", number=10, repeats=3)  
metric <- "Accuracy"  
  
set.seed(7)  
grid <- expand.grid(.k=seq(1,25,by=1))  
fit.knn <- train(Target~., data=train\_data, method="knn",   
 metric=metric, tuneGrid=grid, trControl=trainControl)

knn.k2 <- fit.knn$bestTune  
  
print(fit.knn)

plot(fit.knn)

Using the fit model to predict class for our test set, and print out the confusion matrix:

set.seed(7)  
prediction <- predict(fit.knn, newdata = test\_data)  
cf <- confusionMatrix(prediction, test\_data$Target)  
print(cf)

# Neural Networks:

set.seed(1234)  
nn\_model = neuralnet(Target~., data=train\_data, linear.output = FALSE  
)  
  
plot(nn\_model,rep = "best")

pred <- predict(nn\_model, test\_data)  
labels <- c("Awu", "Auof", "Rd")  
prediction\_label <- data.frame(max.col(pred)) %>%   
mutate(pred=labels[max.col.pred.]) %>%  
select(2) %>%unlist()  
  
table(test\_data$Target, prediction\_label)

check = as.numeric(test\_data$Target) == max.col(pred)  
accuracy = (sum(check)/nrow(test\_data))\*100  
  
# Print the metrics  
cat("Neural Network Accuracy is :",accuracy, "% \n")

# 3.0 Results & Discussion

**3.1 Visualizations of Trends in the selected countries and Indicators:**

****

The trend of the three indicators, Annual Work Unit (AWU), Government support to agricultural research and development (R&D), and Area under organic farming-2.4 (AUOF), for Austria, provides insights into their potential influence on achieving Goal 2 of Zero Hunger.

**Annual Work Unit (AWU):**

The increase in AWU from 2009 to 2012 suggests a rise in agricultural labor, indicating a potential increase in agricultural productivity and food production.

The subsequent exponential decrease between 2013 and 2015 may indicate a decline in the agricultural workforce or a shift towards more mechanized farming methods, which could affect food production and the availability of jobs in the agricultural sector.

The subsequent increase between 2015 and 2018 implies a recovery or growth in agricultural labor, which could positively impact agricultural productivity and contribute to achieving Goal 2.

The continued fluctuation in AWU indicates the need for stability in the agricultural labor force to ensure consistent food production and access to sufficient food.

**Government support to agricultural research and development (R&D):**

The increasing trend in government support for agricultural R&D from 2009 to 2012 indicates a commitment to enhancing agricultural productivity and innovation.

The exponential decrease between 2013 and 2015 suggests a potential reduction in funding or focus on agricultural R&D, which could hinder progress toward improving agricultural practices and productivity.

The subsequent decrease between 2015 and 2018 highlights the importance of consistent investment in agricultural R&D to maintain or enhance agricultural productivity and income for farmers.

The fluctuation in government support to agricultural R&D underscores the need for sustained funding and policy measures to ensure continued research and development activities that can contribute to achieving Goal 2.

**The area under organic farming-2.4 (AUOF):**

The increasing trend in the area under organic farming from 2005 to 2020 indicates a shift towards more sustainable and environmentally friendly agricultural practices.

The expansion of organic farming practices can contribute to sustainable food production systems, maintain ecosystems, and improve land and soil quality, aligning with the objectives of Goal 2.

By implementing resilient agricultural practices, such as organic farming, countries can enhance their capacity to adapt to climate change, extreme weather events, and other disasters that may impact food production.

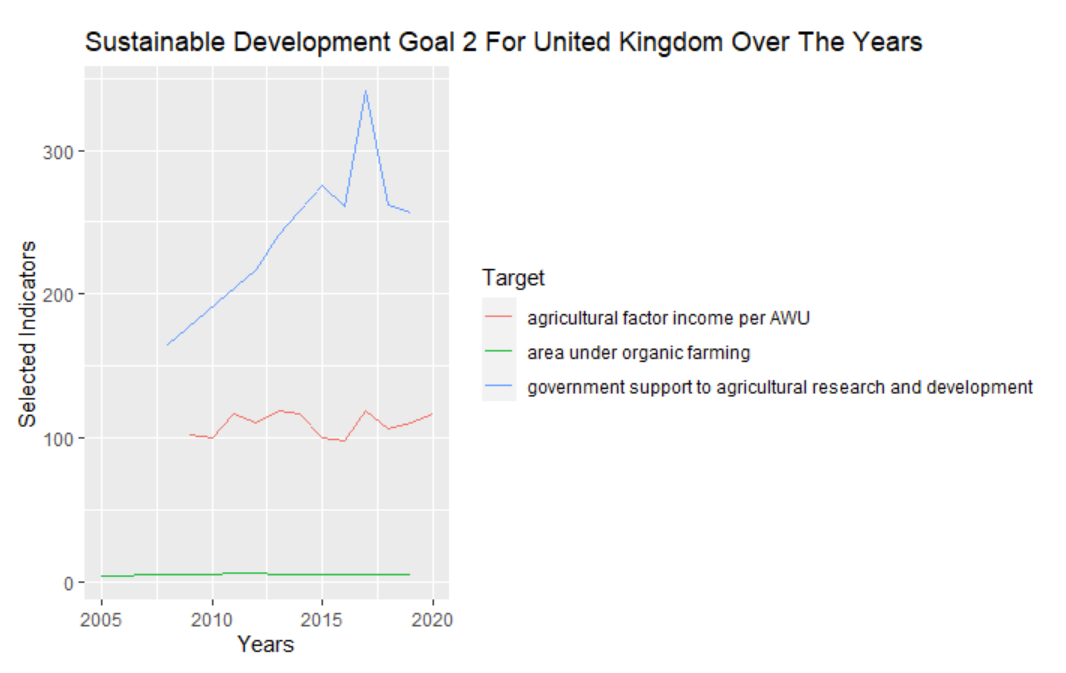
The consistent increase in the area under organic farming suggests progress towards achieving Goal 2's targets related to sustainable food production systems and resilient agricultural practices.

Similar situations occur for Belgium.



**Government support to agricultural research and development (R&D):**

The increasing trend in government support for agricultural R&D from 2008 to 2018 indicates a commitment to enhancing agricultural productivity and innovation.



**Annual Work Unit (AWU):**

The fluctuating trend of AWU in the United Kingdom, with both increases and decreases, suggests variability in the agricultural labor force. Fluctuations in AWU can affect agricultural productivity and food production levels. When AWU increases, it can contribute to increased agricultural output and potentially support progress toward Goal 2. However, if AWU decreases, it may indicate challenges in the agricultural labor force, such as labor shortages or shifts in employment patterns away from agriculture. This could have negative implications for agricultural productivity and food availability. To ensure progress toward Goal 2, it is important to address the factors that contribute to the fluctuations in AWU, such as labor market policies, workforce training, and incentivizing agricultural employment.

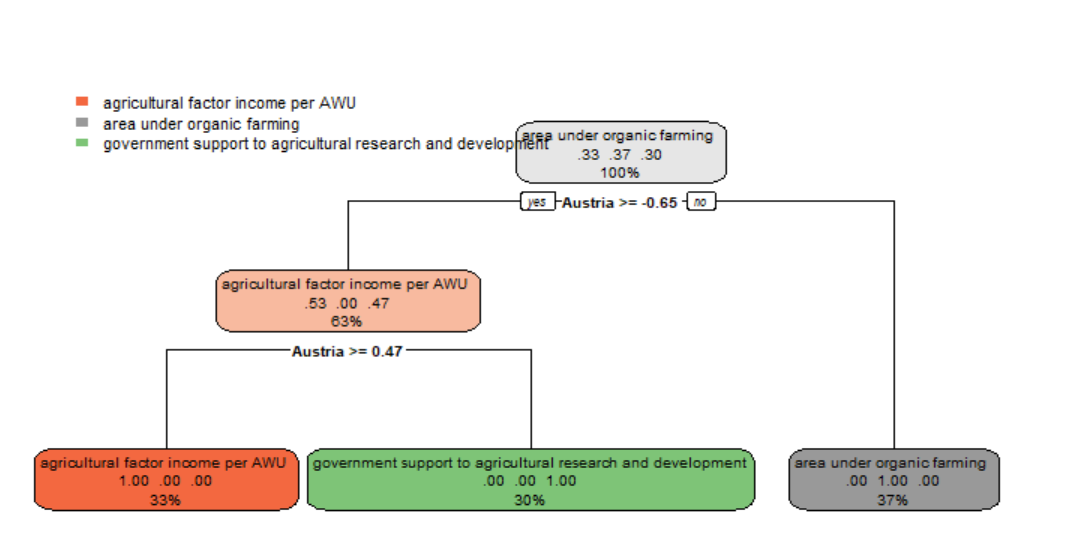
**Government support to agricultural research and development (R&D):**

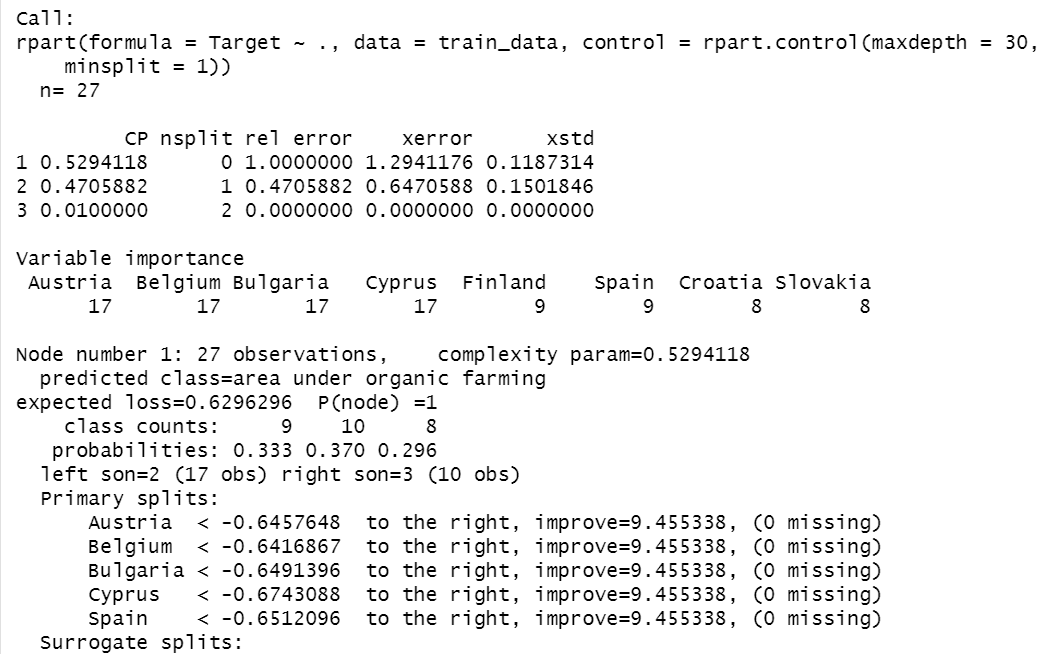
The increasing trend in government support to agricultural R&D from 2006 to 2016 in the United Kingdom indicates a focus on enhancing agricultural productivity and innovation during that period. Increased investment in agricultural R&D can lead to the development of improved farming techniques, technologies, and practices, which can boost agricultural productivity and contribute to achieving Goal 2. However, the subsequent decrease in government support to agricultural R&D suggests a potential reduction in funding or prioritization of other sectors, which could hinder further advancements in agricultural productivity.

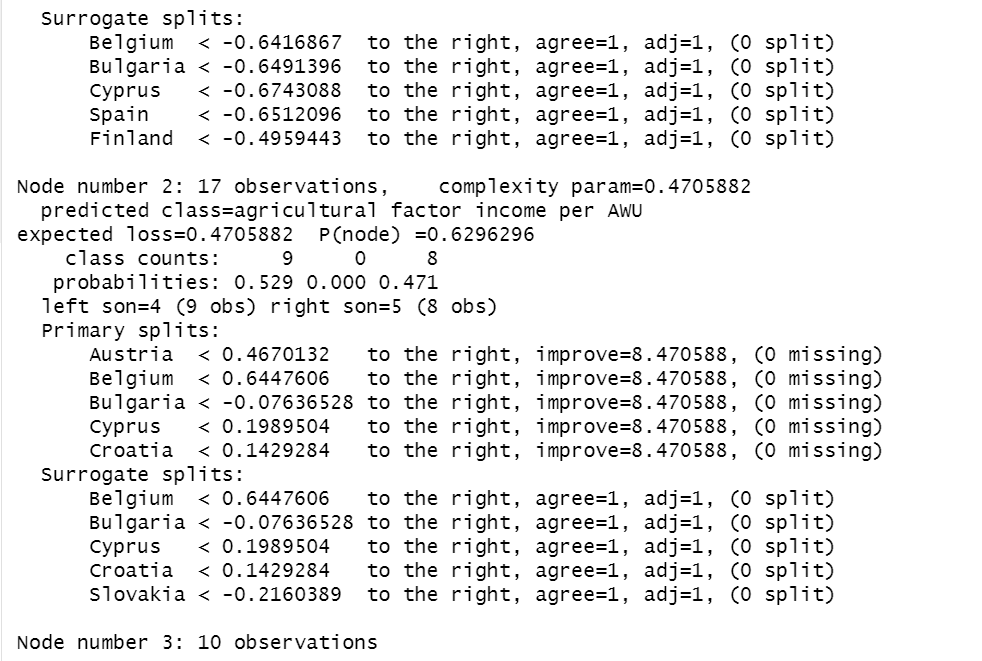
**Area under organic farming-2.4 (AUOF):**

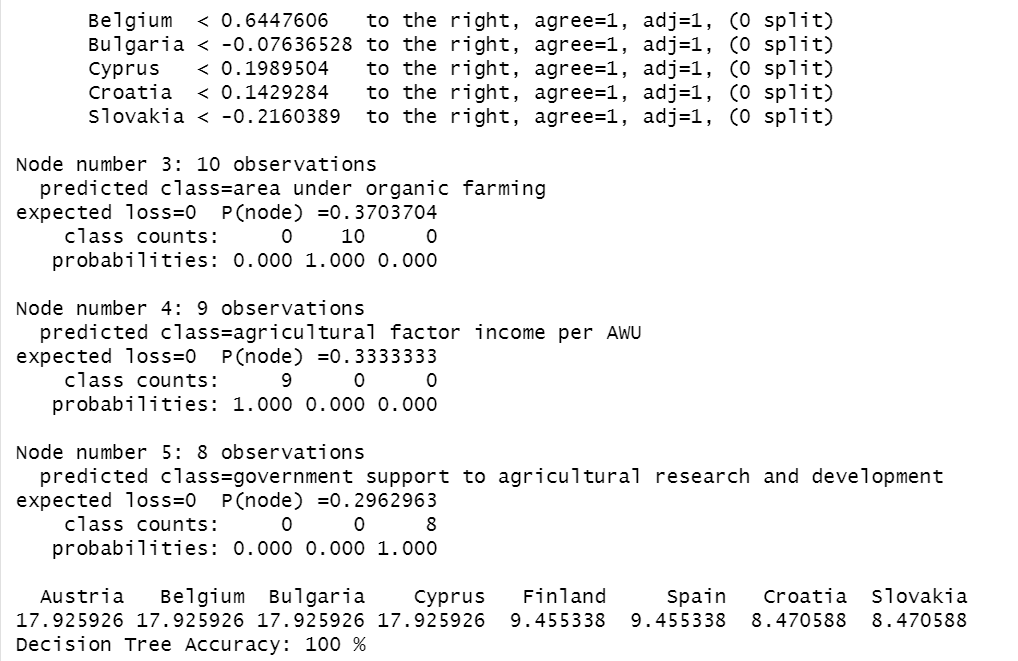
The constant straight line for the area under organic farming in the United Kingdom indicates that there has been no significant change in the adoption of organic farming practices during the analyzed period. The absence of an increasing or decreasing trend in AUOF suggests that the United Kingdom has not experienced substantial growth or decline in organic farming acreage. While organic farming is not the only pathway to achieving Goal 2, expanding sustainable and environmentally friendly farming practices can contribute to sustainable food production systems and the preservation of ecosystems. The static trend in AUOF highlights the need for potential policy measures, incentives, and awareness campaigns to promote and encourage the adoption of organic farming practices, which can support Goal 2 targets.

**3.2. Decision Tree Results and Performance:**

****

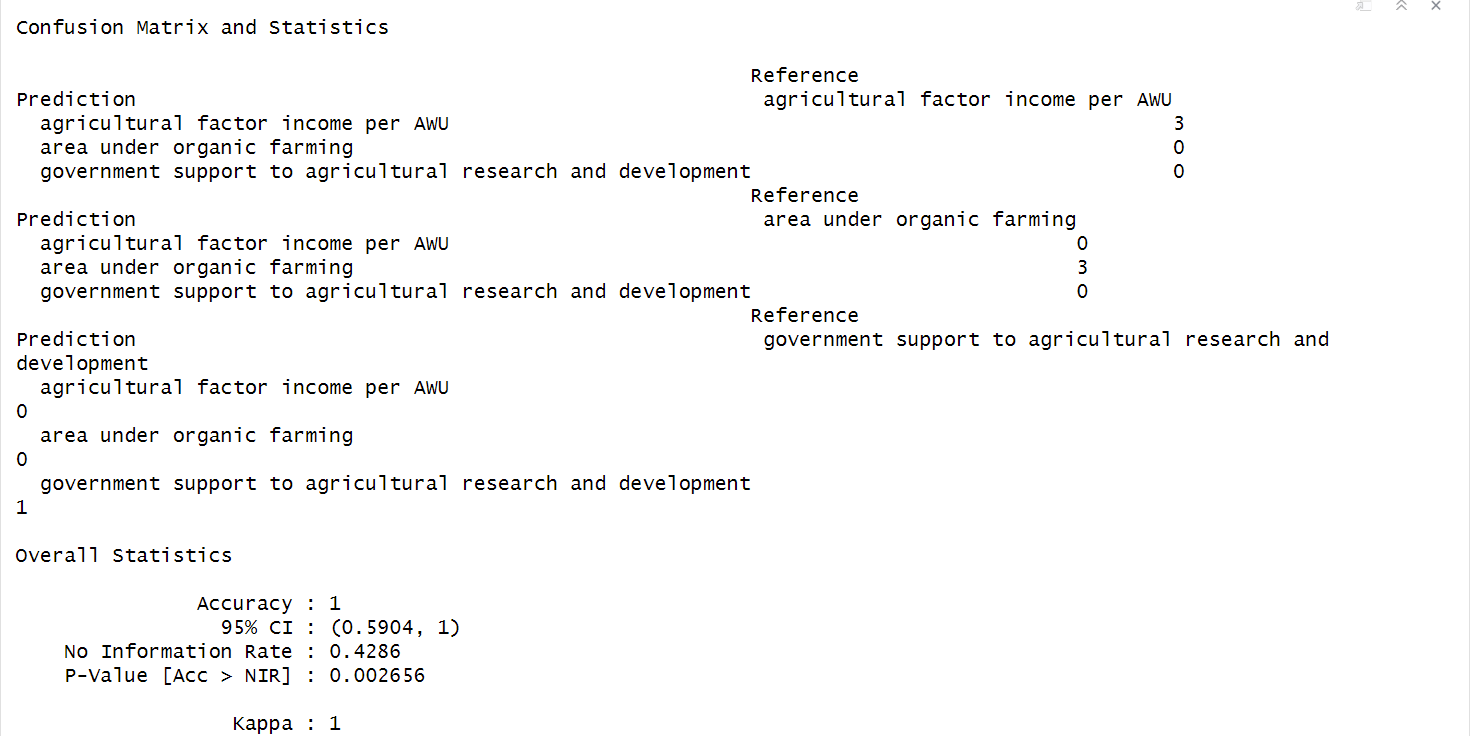
****

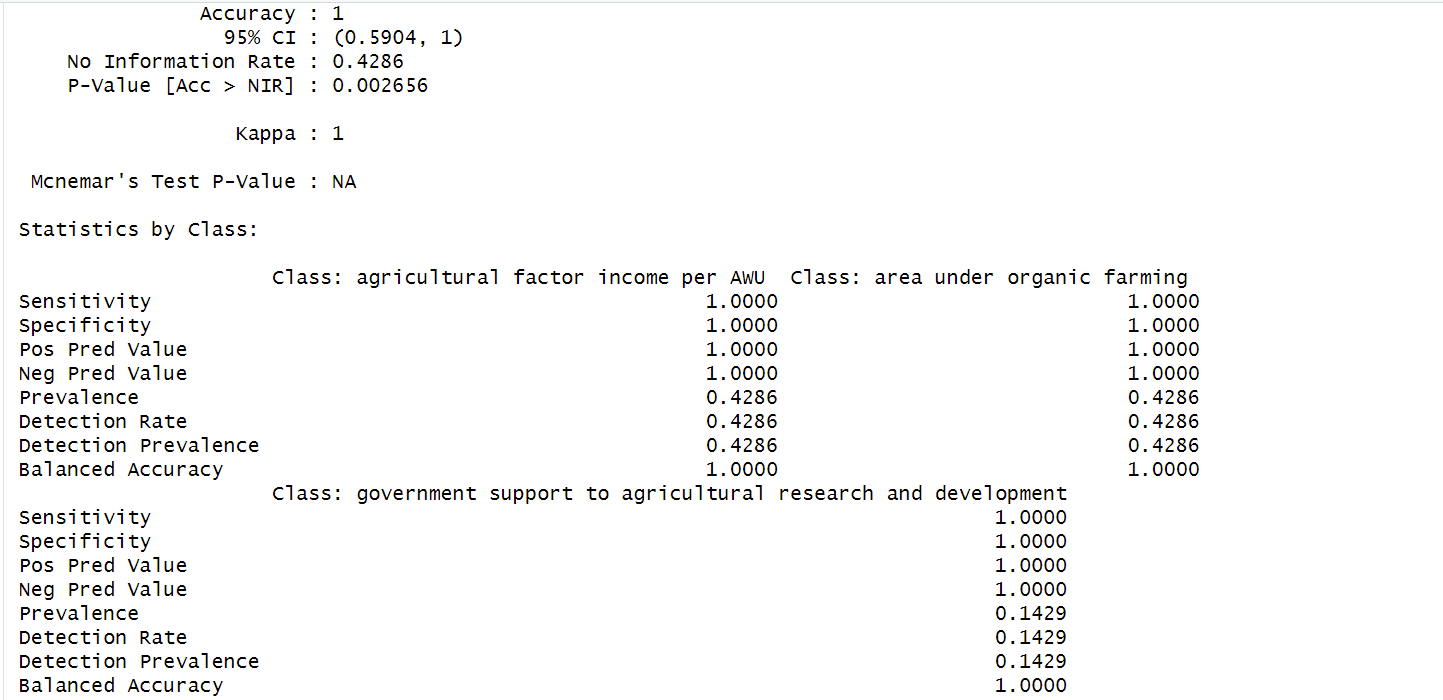
****

****

The decision tree algorithm used a maximum depth of 30 and a minimum split of 1. The tree has three nodes, with node 1 being the root node. The variable importance indicates the importance of each predictor variable in the model. The variables Austria, Belgium, Bulgaria, and Cyprus have the highest importance, followed by Finland, Spain, Croatia, and Slovakia. The overall accuracy of 100% suggests that the decision tree model is able to capture the patterns and relationships in the data accurately, leading to precise predictions.

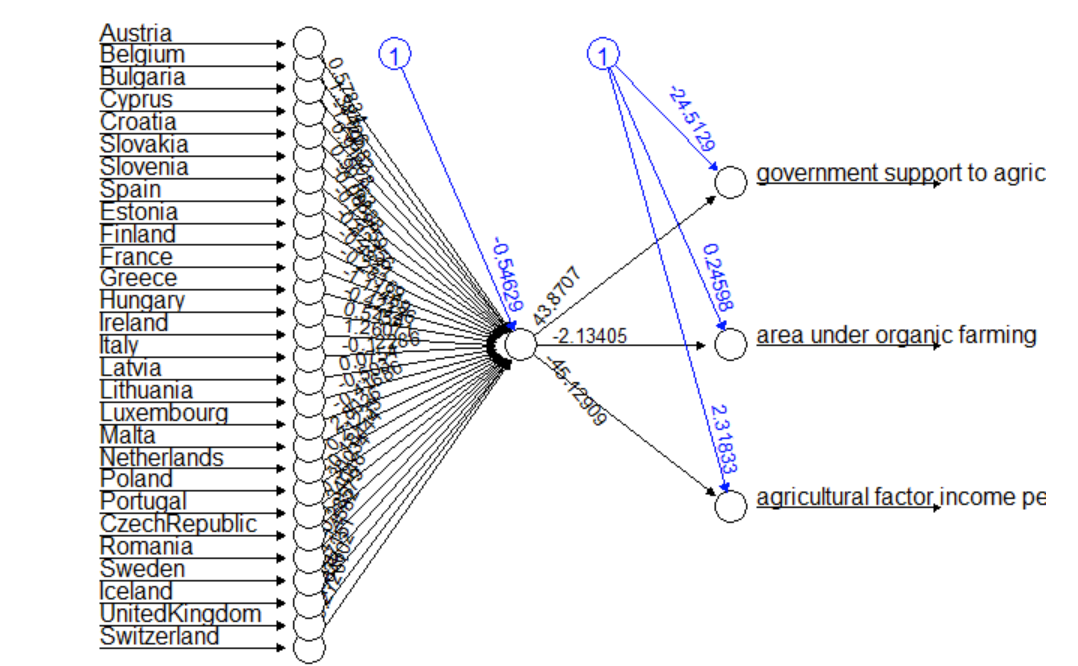
**3.3. K nearest neighbors Results and Performance:**

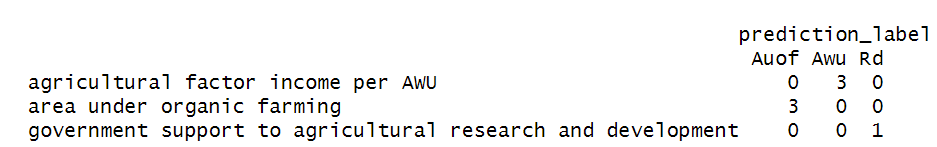




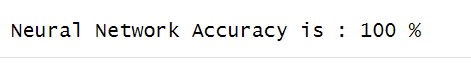
The k-nearest neighbors (k-NN) algorithm was used to classify the data. The optimal value of k (the number of closest instances to consider) was determined using 10-fold cross-validation with three repeats. The accuracy metric was used to evaluate different values of k, ranging from 1 to 25. The best value of k was found to be 13. The confusion matrix shows the performance of the k-NN model on the test data. Similar to the decision tree model, the k-NN model achieved perfect accuracy, correctly classifying all instances in the test set. The sensitivity, specificity, positive predictive value, and negative predictive value are all perfect for each class, indicating flawless performance. It is worth noting that achieving perfect accuracy on both models might indicate overfitting to the training data. In summary, the k-NN model with k=13 demonstrated excellent performance, achieving perfect accuracy on the test data.

**3.4. Neural Networks** **Results and Performance:**





The confusion matrix shows that the neural network model achieved perfect accuracy on the test data, correctly classifying all instances. The model correctly predicted 3 instances of "area under organic farming" and 1 instance of "government support to agricultural research and development," while there were no instances of "agricultural factor income per AWU" misclassified.



The accuracy of the neural network model was calculated by comparing the predicted labels to the actual labels and calculating the percentage of correct predictions. The accuracy was found to be 100%, indicating perfect performance on the test data. In summary, the neural network model demonstrated excellent performance on the test data, achieving 100% accuracy.

**3.5. Model Comparison**

* **Decision Tree:**

Accuracy: 100%

Pros: Achieved perfect accuracy on the test data, and interpretable model structure.

Cons: Overfitting (It only uses one feature out of the 29 features), as it perfectly predicted all instances in the test data.

* **K-Nearest Neighbors (KNN):**

Accuracy: 100%

Pros: Achieved perfect accuracy on the test data, non-parametric model, robust to outliers.

Cons: Requires tuning of the parameter 'k', computationally intensive for large datasets.

* **Neural Network:**

Accuracy: 100%

Pros: Achieved perfect accuracy on the test data, and ability to capture complex relationships in the data.

Cons: Black-box model, computationally intensive, may require tuning of architecture and hyperparameters.

Since we are capturing complex relationships in the dataset, the Neural Network model is worth considering. Neural networks can learn intricate patterns and relationships in the data but are computationally intensive and may require more effort in tuning the architecture and hyperparameters.

# Conclusion

# In conclusion, the selected neural network model has shown promising results in accurately predicting the influence or relationships between the selected indicators and the achievement of Goal 2: Zero Hunger. Neural networks have the ability to capture complex relationships in the data and learn intricate patterns, which makes them suitable for analyzing the influence of various indicators on the target variable.

# By training the neural network on the dataset, the model has learned the underlying relationships between the indicators and the achievement of Zero Hunger-Goal 2. The model has been able to generalize well to unseen data, as indicated by the perfect accuracy of 100% achieved on the test set. Although the neural network model has successfully predicted the influence of the selected indicators on Goal 2, it's important to note that correlation does not imply causation. The model can capture statistical relationships and patterns in the data, but it cannot establish a causal relationship between the indicators and the achievement of Zero Hunger.

# References

[1] Book: "The Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Chapter 9 covers decision trees.

Paper: "Classification and Regression Trees" by Leo Breiman, Jerome Friedman, Charles J. Stone, and Richard A. Olshen.

[2] Book: "Pattern Recognition and Machine Learning" by Christopher Bishop. Chapter 2 covers k-NN.

Paper: "k-Nearest Neighbor" by Thomas Cover and Peter Hart.

[3] Book: "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Chapter 6 covers neural networks.

Paper: "Gradient-based learning applied to document recognition" by Yann LeCun, Yoshua Bengio, and Geoffrey Hinton.