

# Application of Machine Learning Techniques to Animal Health Data: A Scoping Project



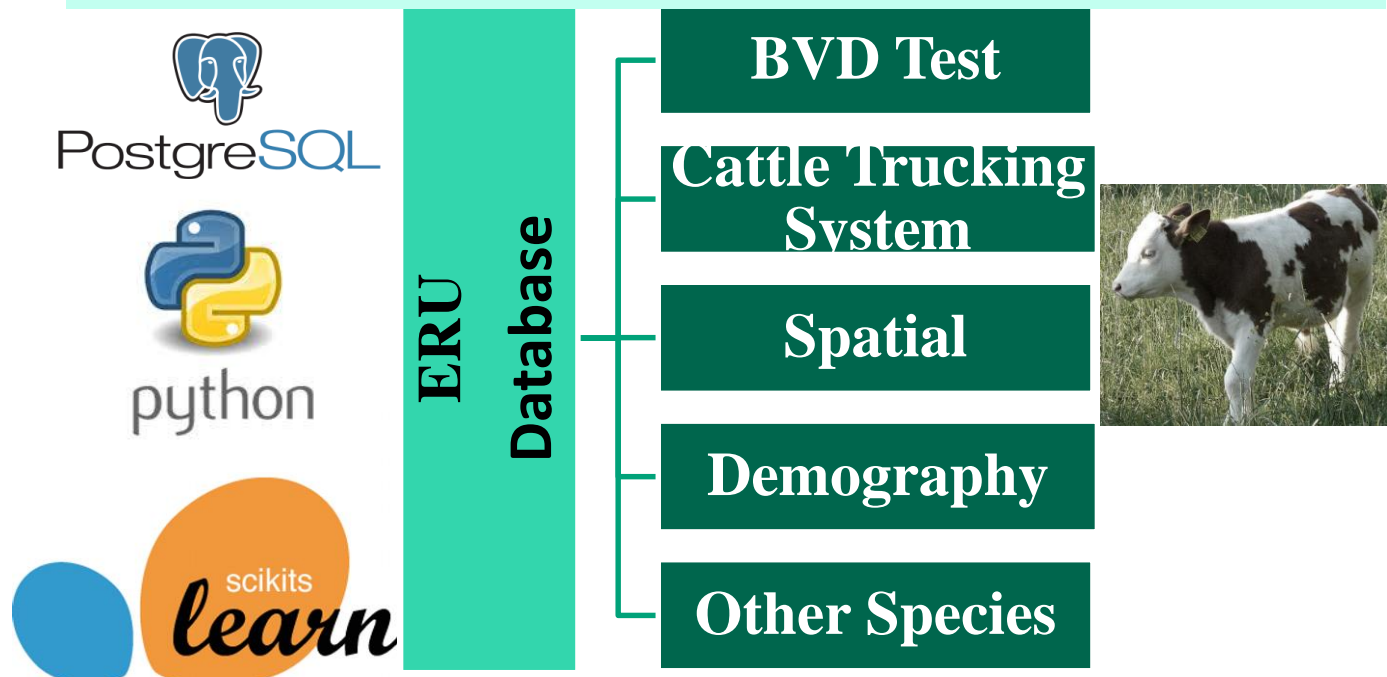
Frank Owusu Mensah  
MSc in Big Data

## Introduction

**Problem Statement:** Bovine Viral Diarrhea (BVD) is an endemic disease of cattle that has huge health and economic impacts. Since 2010, the Government has embarked on a project to eradicate it from Scotland. While some farms are now BVD free, other are not. Yet still others keep swinging their BVD status.

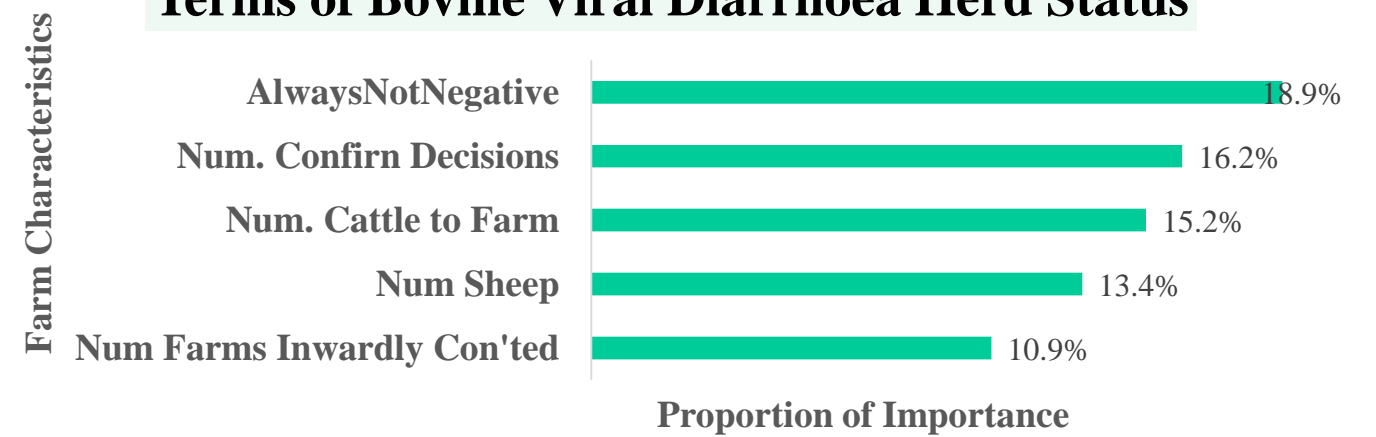
**Goal:** To employ Machine Learning techniques to identify the characteristics of farms that are not yet free from BVD using the datasets of Epidemiology Research Unit (ERU).

## Data Source & Software



## Results and Discussion

### Top Five Important Farm Characteristics in Terms of Bovine Viral Diarrhoea Herd Status

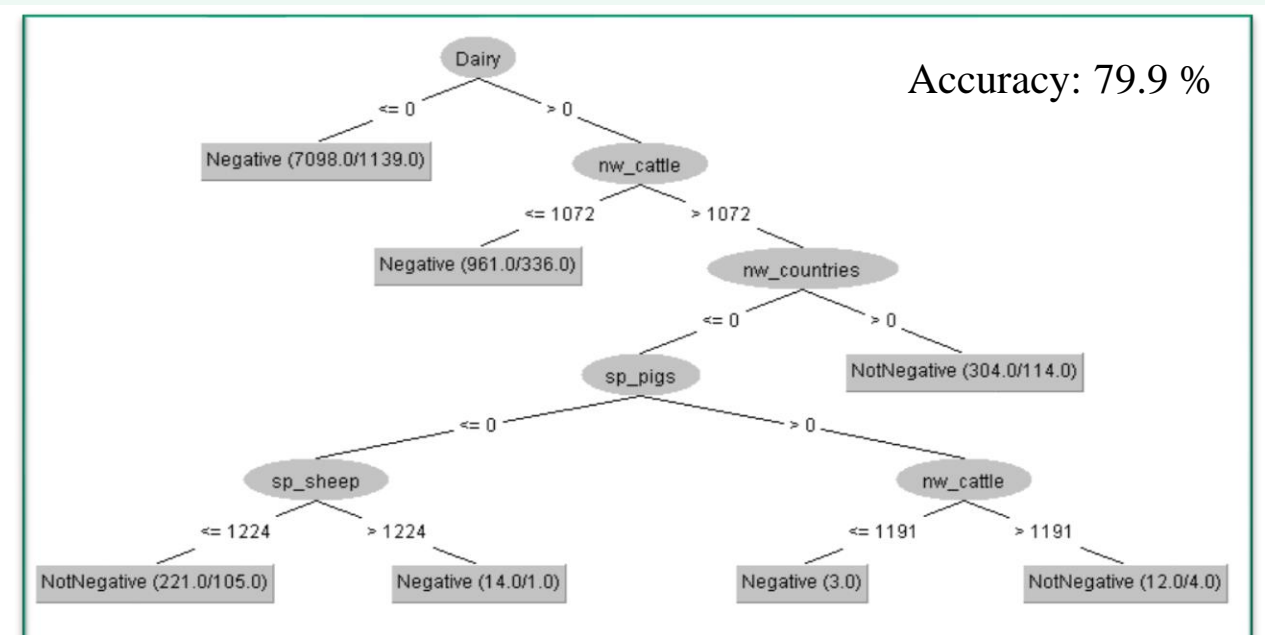


With performance of  $84.4 \pm 1.0\%$ , the most important feature that determines farm's BVD status is whether or not a farm's status has always been **NotNegative**.

Also important are the **in-degree** centrality in terms of number of cattle **moved** to a farm and number of farms inwardly **connected** to a farm.

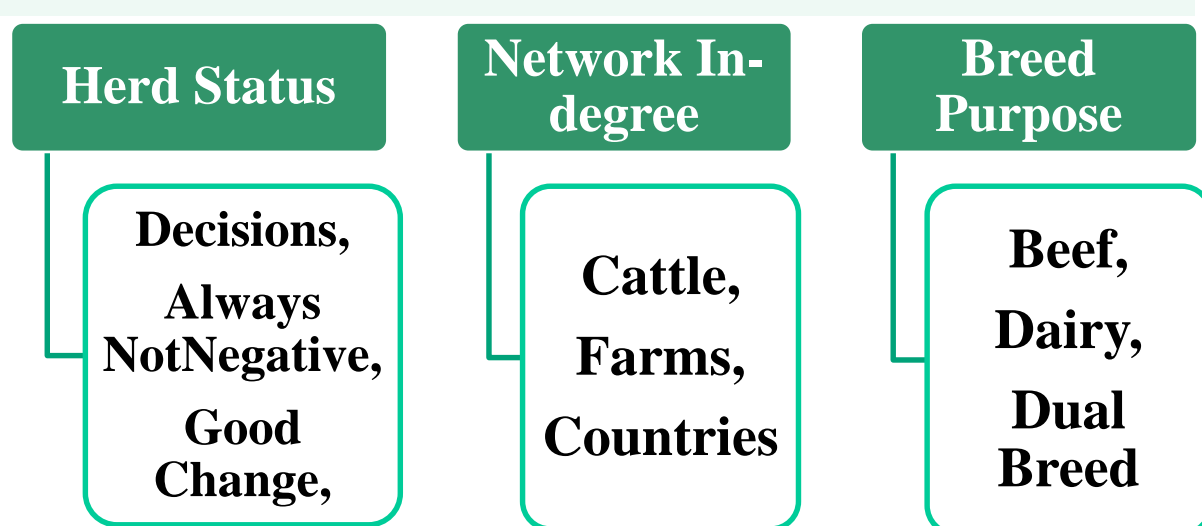
Furthermore, the tree below shows that **Dairy** cattle, number of **countries** a farm has imported cattle from and presence of **pigs** influence a farm's BVD status.

### Decision Tree Without Herd Status (Weka)

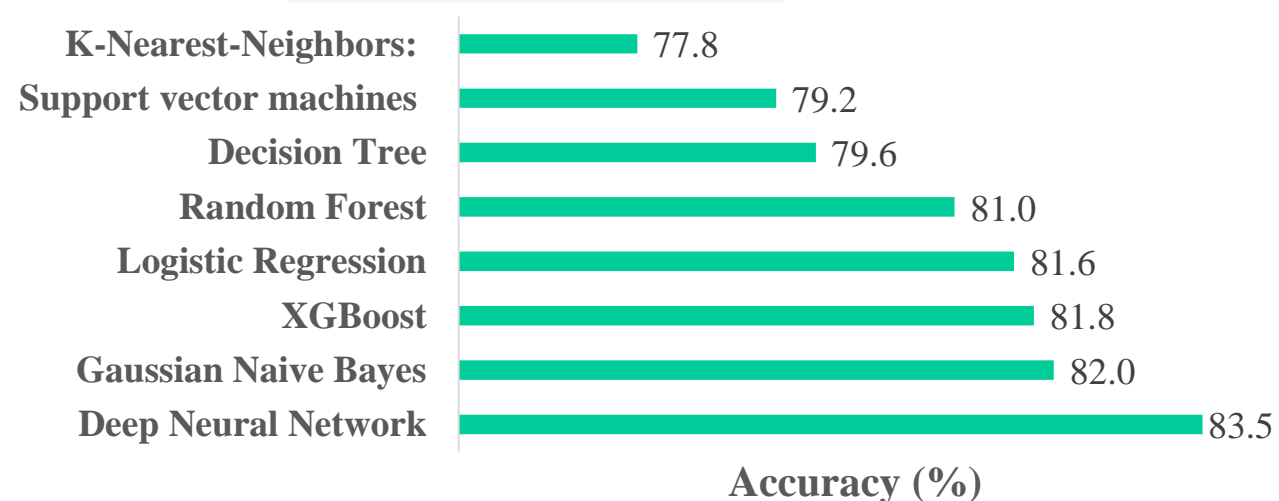


## Methodology

### 1. Farm Characteristics Extraction



### 2. Model Selection



## Conclusion and Recommendations

This study has shown that farm characteristics such as Always-NotNegative, in-degree connection with other farms, Dairy cattle and pigs directly affect farm's NotNegative status.

Therefore, rules governing animal movement especially those from outside Scotland must be strengthened and more attention given to farms with Dairy cattle.

Finally, further study is needed to ascertain whether or not pigs serve as BVD reservoir for cattle.