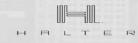




Halter Project

- Halter is a pioneering company leveraging technology for dairy farmers and operates farms throughout New Zealand.
- Halter provides collars with a specialised app to collect sensor data from cows on farms.
- Sensor data is transformed into valuable behavioural insights through advanced modelling.
- The combined analysis of sensor and behaviour data enables effective monitoring and detection of cows' health conditions.
- Train dataset: 10,181 cows from 150 farms, 5 metadata, 12 sensor and 12 behaviour fs
- Test dataset: 409 cows from 25 farms, 12 sensor variables and 12 behaviour variables in both datasets.



Project Purpose

Objectives

- identify significant correlations within the variables
- explore seasonal sickness
- characterise the recovery process of cows following sickness
- establish patterns to estimate duration of a cow's sickness

Benefits to Farmers

- managing the health of their cows
- preventing mild conditions from worsening and
- minimising the spread of sickness to other cows
- safeguarding the well-being of cows, restoring full productivity and preserving valuable assets







Train Dataset

Metadata

UTC Timestamp

Cattle ID

Farm ID

Sickness Type

Hours Since Sickness

Behaviour Features

Grazing

Mob Median Grazing

Mob Std Grazing

Farm Average Grazing

Resting

Mob Median Resting

Mob Std Resting

Farm Average Resting

Rumination

Mob Median Rumination

Mob Std Rumination

Farm Average Rumination

Sensor Features

ODBA Average

ODBA Std

Farm Average ODBA Average

Farm Average ODBA Std

Pitch Average

Pitch Std

Farm Average Pitch Average

Farm Average Pitch Std

Roll Average

Roll Std

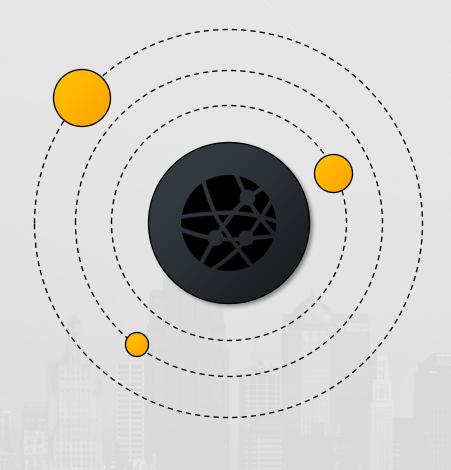
Farm Average Roll Average

Farm Average Roll Std





Data Wrangling





Focus on up to 14 days of pre- and post-health/sickness



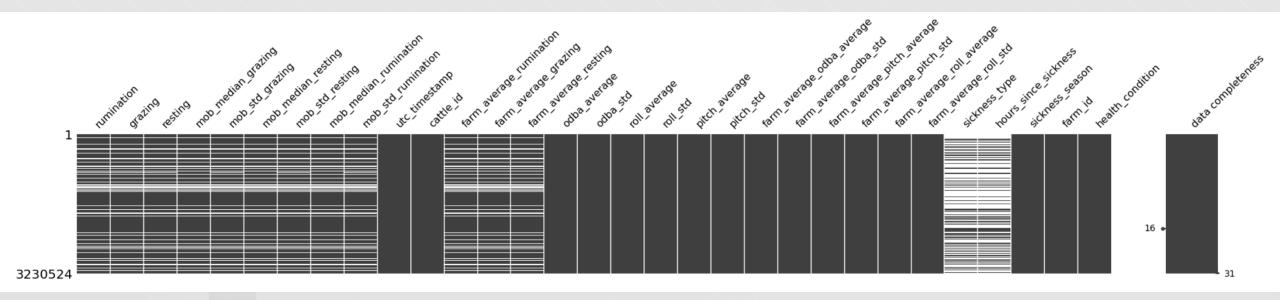
Find a target timestamp to select up to 14 days of pre- and post-health/sickness

Look at the hours since sickness column per cow:

- if all blanks: target timestamp=the earliest timestamp
- > at least one value, find 0 or the absolute value closest to 0:
- hours since sickness=0, target timestamp on the same row;
- hours since sickness>0, target timestamp-(hours since sickness);
- hours since sickness<0, target timestamp+(-hours since sickness).



Missing Data

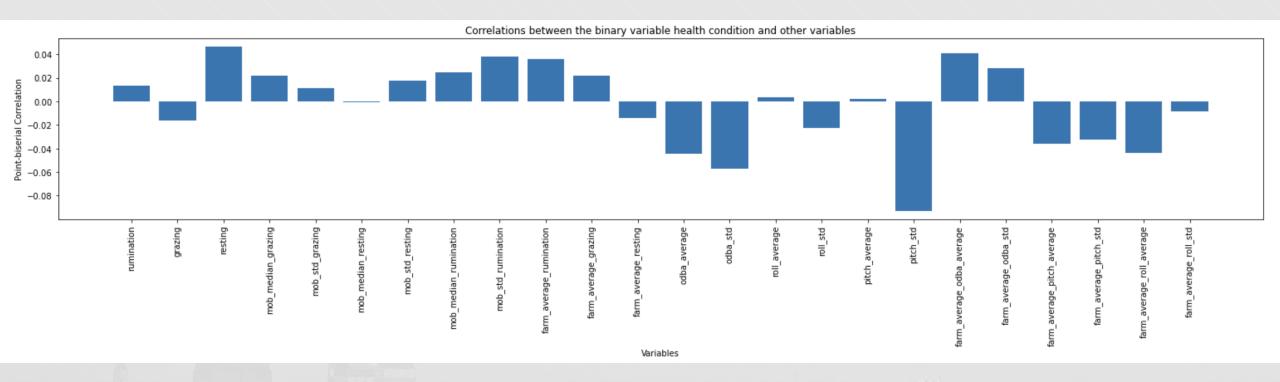


- 31 columns with 3230524 rows
- Full data: 17 columns
- Missing data: 14 columns, may be caused by sickness type and hours since sickness, which means healthy cows





Point-biserial Correlation



- **Highly positively correlated to health condition**: resting, mob std rumination, farm average rumination, mob median rumination
- Highly negatively correlated to health condition: pitch std, odba std, farm average roll average, odba average



Process Flow





Feature Engineering

1 Adding Health Condition

Health_Condition

- 0=health
- 1=sick

2 Adding Seasonal Variable

Sickness_Season

- Spring = Sep-Nov
- Summer= Dec-Feb
- Autumn= Mar-May
- Winter = Jun-Aug

3 Adjusting Relative to Farm/Mob

Adding 18 new variables



Behavior Features
Relative to Farm/Mob
(9)



Sensor Features
Relative to Farm/Mob
(6)



Mob Relative to Farm

(3)

- Grazing Farm_Avg
- Grazing Mob Med
- Grazing Mob_Std
- Resting Farm Avg
- Resting Mob Med
- Resting Mob_Std
- Rumitating Farm Avg
- · Rumitating Mob Med
- Rumitating Mob_Std

- OBDA Farm_Avg
- OBDA Farm_Std
- Pitch Farm Avg
- Pitch Farm Std
- Roll Farm Avg
- Roll Farm Std

- Graze: Mob_med Farm_Avg
- Resting: Mob med Farm Avg
- Rumitate: Mob_med Farm_Avg





Methodologies









Decision Tree

Gradient **Boosting**

Logistic Regression

Neural **Network**

Description

- a simple yet powerful supervised learning model used for classification and regression tasks
- an ensemble learning technique that combines multiple decision trees
- a widely-used statistical model to evaluate the relationship between input features and the probability of an event
- particularly deep learning model, excels at identifying complex patterns and relationships within the data

- **Advantages**
- Highly interpretable
- Highly accurate

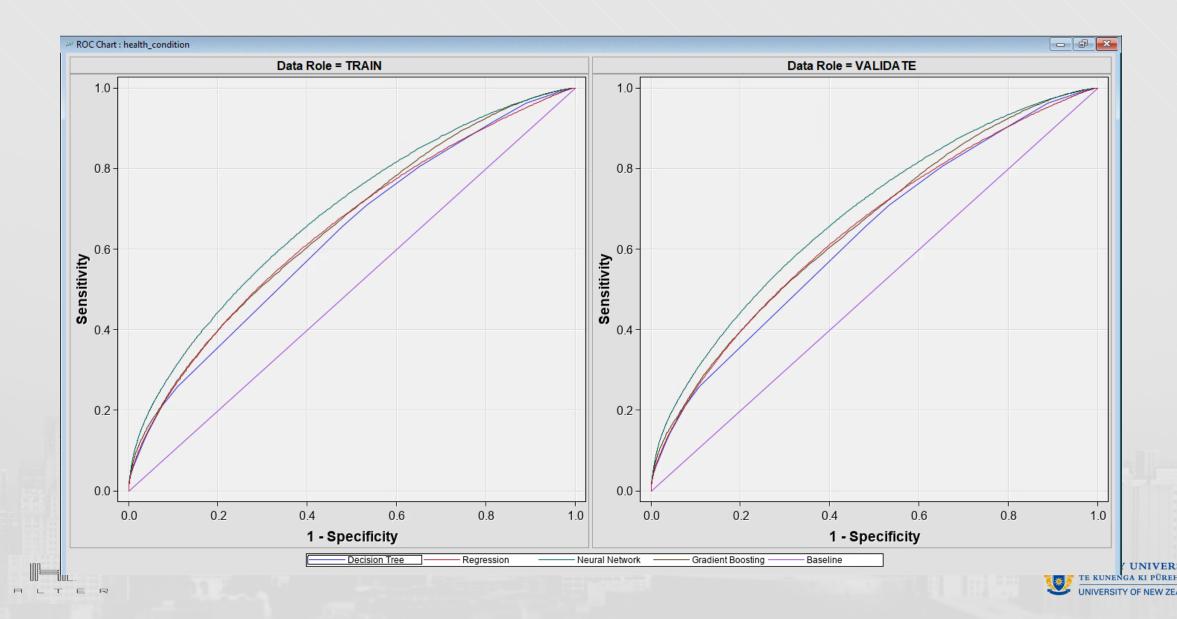
- Efficient
- Interpretable

Highly flexible

- Prone to overfitting
- Less interpretable
- May not perform well with highdimensional or highly correlated features
- Less interpretable
- Computationally intensive MASSEY UNIVERSITY



Model Performance



Evaluation Metrics





Recall

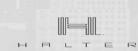
Specificity



Health Cows Not Alerted

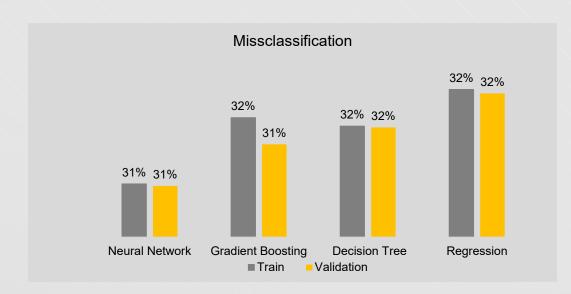
Total Sick Cows

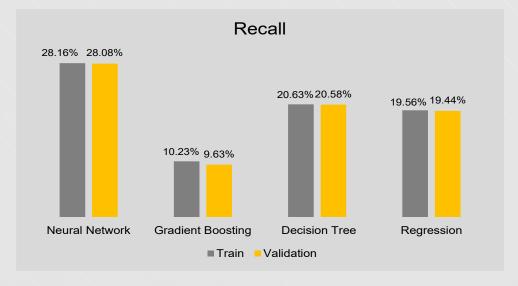
Total Healthy Cows

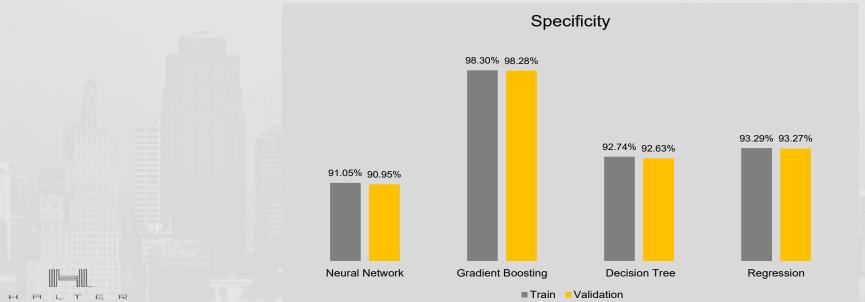


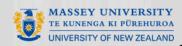


Model Performance

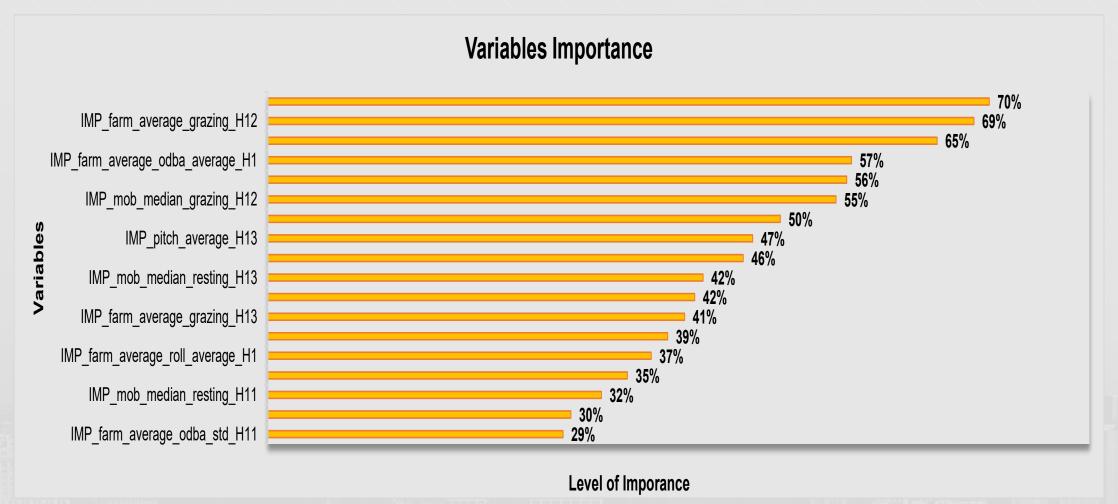








Key Variables of Neural Network





Prediction Score of Neural Network

Recall (percent of sick cows detected, relative to when it was recorded):

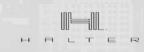
2 weeks before / after: 79%

2 weeks before only: 77%

Precision:

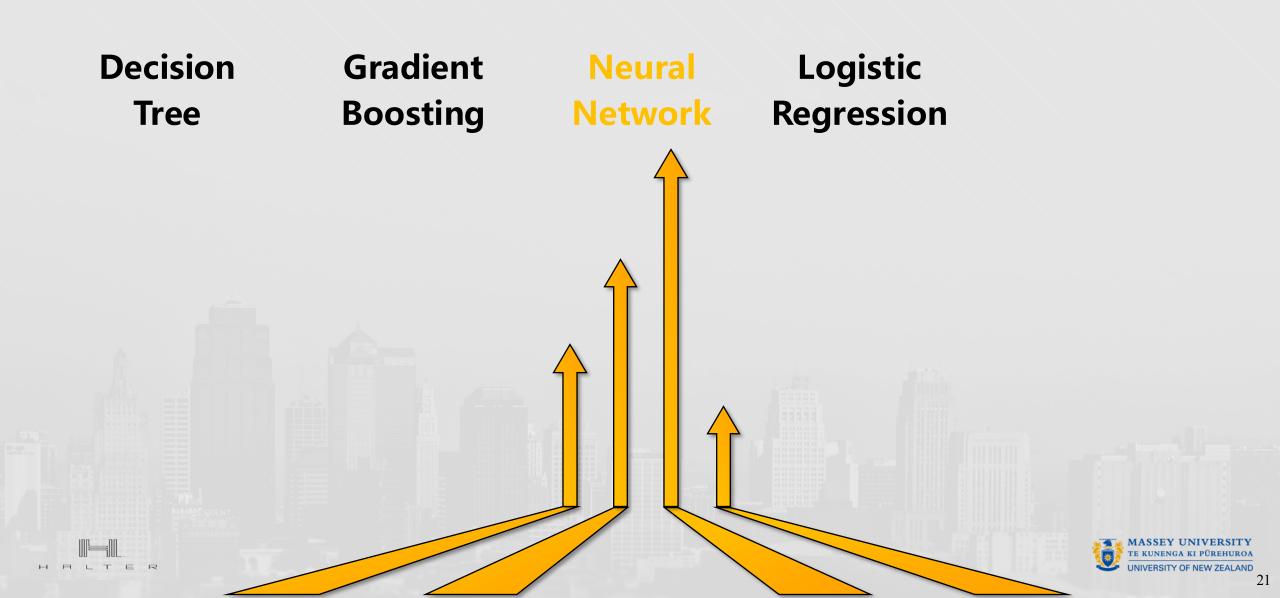
Alerted cows which were sick: 98%

Alerts within 2 weeks of a sickness: 59%





Conclusion



Further Study



- Adding new variables: 18 variables adjusting relatives to farm or mob
- Recovery from sickness: health condition turn from 1 (sick) to 0 (healthy)
- **Sickness duration:** from the time when cow's health condition is 1 to the time when cow's health condition is 0

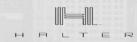


Model improvement



For further study:

- health condition as a binary variable, **long short-term memory networks**
- sickness type as a categorical variable, naïve baye and nominal model





Key Variables

(Adding new variables)

