

1. Main Goals:

Our main goal is to predict a horse's survival outcome—died, euthanized, or survived—to identify those in prime racing condition. Concurrently, we aim to pinpoint key health variables that most influence this outcome, helping prioritize care for optimal horse well-being.

Understanding these factors is crucial for several reasons:

- It can help veterinarians and horse owners make more informed decisions about the health and well-being of the animals.
- It can be particularly beneficial for the horse racing industry to know which horses are most fit to race, thereby reducing the risks of fatal injuries.
- It also has implications for animal welfare, as identifying health issues early can lead to better treatment options.

2. Models and Methods (Supervised):

- K-Nearest Neighbors (KNN)
 - How it Works: Given a new observation, KNN identifies 'K' closest data points in the feature space and assigns a label based on the majority class of these neighbors.
 - Features: Age, pulse, respiratory_rate, and total_protein among others.
 - Parameters: Number of neighbors (K), Distance Metric (Euclidean, Manhattan, etc.)
- Support Vector Machines (SVM)
 - How it Works: SVM tries to find a hyperplane that best divides a dataset into classes.
 - Features: Same as KNN.
 - Parameters: Kernel (Linear, Polynomial, Radial), C (Regularization Parameter), Gamma.
- Performance Comparison:
 - Confusion Matrix: To evaluate the True Positives, True Negatives, False Positives, and False Negatives.
 - Cross-Validation: To assess how the models generalize to an independent dataset.
 - Accuracy: Overall correctness of the model.

3. Workload Distribution:

MATH 4323

10/30/23

Armando Espinoza

Derick Nguyen

Cong Nguyen

Lawrence Castacio

Rachel Collier

- Armando, Derick, and Rachel: Will focus on implementing and testing the KNN model. They will also work on data pre-processing and feature selection for KNN.
- Cong & Lawrence: Will work on the SVM model, fine-tuning the parameters, and testing its efficacy. They will also be responsible for the performance comparison of the two models.