MATH 4323
09/29/23
Armando Espinzoa
Derick Nguyen
Cong Nguyen
Lawrence Castacio
Rachel Collier

1.

   a. **Data source:** Walter Reade, Ashley Chow. (2023). Predict Health Outcomes of Horses. Kaggle.

   https://kaggle.com/competitions/playground-series-s3e22

   b. **Inspiration:** We are interested in horse racing and would like to know which horses would be in the best condition to race.

2. **Train data set:** 1235 rows and 29 columns. **Test data set:** 824 rows and 28 columns

3.

Description

The *Horse* data frame has 2059 rows and 29 columns.

Format:

Id: Horse identification

surgery: Indicates whether the horse has had surgery ("no" or "yes")

age: Indicates whether the horse Is an adult or young

hospital_number: Id for the hospital the horse is held at

rectal_temp: rectal temperature of the horse which is measured in Celsius

pulse: Heart rate of the horse which is measured in beats/minute

respiratory_rate: rate of respiration of the horse measured in beats per minute.

temp_of_extremities: temperature reading of the horse's extremities, categorized as either "None", "cold", "cool", "normal", or "warm".

peripheral_pulse: Heart rate of the horse in peripheral points of the body, categorized as either "reduced", "normal", or "other".

mucous_membrane: the color of the horse's mucus membrane (normal_pink, bright_red, etc.).

capillary_refill_time: The time required for return of color after application of pressure onto a body part i.e. the time it takes for a body part to regain oxygen rich blood (less_3_sec, more_3_sec, etc.).

pain: The pain level of the horse categorized as alert, no pain, depressed, intermittent mild pain, intermittent severe pain, or continuous severe pain

peristalsis: description of gut motility (hypomotile, hypermotile, normal, absent, None).

abdominal_distention: the level of abdominal distention observed (slight, moderate, etc.).

nasogastric_tube: The presence of gas in the nasogastric tube (none, slight, or significant).

nasogastric_reflux: The amount of nasogastric reflux (none, < 1 liter, or > 1 liter)

nasogastric_reflux_ph: The pH value of the nasogastric reflux

rectal_exam_feces: The amount of feces observed during rectal examination (normal, decreased, etc.).

abdomen: An assessment of a horse's abdomen (distend_small, normal, etc.)

packed_cell_volume: The number of red cells by volume in blood

total_protein: The level of total protein in the blood

abdomo_appearance: The appearance of a horse's abdominal fluid (clear, cloudy, serosanguinous, etc.).

abdomo_protein: The level of protein level in abdominal fluid, if present

surgical_lesion: Indicates the presence of a lesion ("yes" or "no")

lesion_1: The area in which a lesion is found

lesion_2: The type of lesion

lesion_3: The subtype of the lesion

cp_data: Indicates whether pathology data was present ("yes" or "no")

outcome: 3 possible values - died, euthanized or survived that indicate the status of a horse