# Horse Colic Data Set

*Download*: Data Folder, Data Set Description

**Abstract**: Well documented attributes; 368 instances with 28 attributes (continuous, discrete, and nominal); 30% missing values

| | | | | | |
|---|---|---|---|---|---|
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 368 | **Area:** | Life |
| **Attribute Characteristics:** | Categorical, Integer, Real | **Number of Attributes:** | 27 | **Date Donated** | 1989-08-06 |
| **Associated Tasks:** | Classification | **Missing Values?** | Yes | **Number of Web Hits:** | 39105 |

## Source:

Creators:

Mary McLeish & Matt Cecile
Department of Computer Science
University of Guelph
Guelph, Ontario, Canada N1G 2W1
mdmcleish '@' water.waterloo.edu

Donor:

Will Taylor (taylor '@' pluto.arc.nasa.gov)

## Data Set Information:

2 data files:
-- horse-colic.data: 300 training instances
-- horse-colic.test: 68 test instances

Possible class attributes: 24 (whether lesion is surgical)
-- others include: 23, 25, 26, and 27

Many Data types: (continuous, discrete, and nominal)

## Attribute Information:

1: surgery?
1 = Yes, it had surgery
2 = It was treated without surgery

2: Age
1 = Adult horse
2 = Young (< 6 months)

3: Hospital Number
- numeric id
- the case number assigned to the horse (may not be unique if the horse is treated > 1 time)

4: rectal temperature
- linear
- in degrees celsius.
- An elevated temp may occur due to infection.
- temperature may be reduced when the animal is in late shock
- normal temp is 37.8
- this parameter will usually change as the problem progresses, eg. may start out normal, then become elevated because of the lesion, passing back through the normal range as the horse goes into shock
5: pulse
- linear
- the heart rate in beats per minute
- is a reflection of the heart condition: 30 -40 is normal for adults
- rare to have a lower than normal rate although athletic horses may have a rate of 20-25
- animals with painful lesions or suffering from circulatory shock may have an elevated heart rate

6: respiratory rate
- linear
- normal rate is 8 to 10
- usefulness is doubtful due to the great fluctuations

7: temperature of extremities
- a subjective indication of peripheral circulation
- possible values:
1 = Normal
2 = Warm
3 = Cool
4 = Cold
- cool to cold extremities indicate possible shock
- hot extremities should correlate with an elevated rectal temp.

8: peripheral pulse
- subjective
- possible values are:
1 = normal
2 = increased
3 = reduced
4 = absent
- normal or increased p.p. are indicative of adequate circulation while reduced or absent indicate poor perfusion

9: mucous membranes
- a subjective measurement of colour
- possible values are:
1 = normal pink
2 = bright pink
3 = pale pink
4 = pale cyanotic
5 = bright red / injected
6 = dark cyanotic
- 1 and 2 probably indicate a normal or slightly increased circulation
- 3 may occur in early shock
- 4 and 6 are indicative of serious circulatory compromise
- 5 is more indicative of a septicemia

10: capillary refill time
- a clinical judgement. The longer the refill, the poorer the circulation
- possible values
1 = < 3 seconds
2 = >= 3 seconds

11: pain - a subjective judgement of the horse's pain level
- possible values:
1 = alert, no pain
2 = depressed
3 = intermittent mild pain
4 = intermittent severe pain
5 = continuous severe pain
- should NOT be treated as a ordered or discrete variable!
- In general, the more painful, the more likely it is to require surgery
- prior treatment of pain may mask the pain level to some extent

12: peristalsis
- an indication of the activity in the horse's gut. As the gut becomes more distended or the horse becomes more toxic, the activity decreases
- possible values:
1 = hypermotile
2 = normal
3 = hypomotile
4 = absent

13: abdominal distension
- An IMPORTANT parameter.
- possible values
1 = none
2 = slight
3 = moderate
4 = severe
- an animal with abdominal distension is likely to be painful and have reduced gut motility.
- a horse with severe abdominal distension is likely to require surgery just tio relieve the pressure

14: nasogastric tube
- this refers to any gas coming out of the tube
- possible values:
1 = none
2 = slight
3 = significant

- a large gas cap in the stomach is likely to give the horse discomfort

15: nasogastric reflux
- possible values
1 = none
2 = > 1 liter
3 = < 1 liter
- the greater amount of reflux, the more likelihood that there is some serious obstruction to the fluid passage from the rest of the intestine

16: nasogastric reflux PH
- linear
- scale is from 0 to 14 with 7 being neutral
- normal values are in the 3 to 4 range

17: rectal examination - feces
- possible values
1 = normal
2 = increased
3 = decreased
4 = absent
- absent feces probably indicates an obstruction

18: abdomen
- possible values
1 = normal
2 = other
3 = firm feces in the large intestine
4 = distended small intestine
5 = distended large intestine
- 3 is probably an obstruction caused by a mechanical impaction and is normally treated medically
- 4 and 5 indicate a surgical lesion

19: packed cell volume
- linear
- the # of red cells by volume in the blood
- normal range is 30 to 50. The level rises as the circulation becomes compromised or as the animal becomes dehydrated.

20: total protein
- linear
- normal values lie in the 6-7.5 (gms/dL) range
- the higher the value the greater the dehydration

21: abdominocentesis appearance
- a needle is put in the horse's abdomen and fluid is obtained from
the abdominal cavity
- possible values:
1 = clear
2 = cloudy
3 = serosanguinous
- normal fluid is clear while cloudy or serosanguinous indicates a compromised gut

22: abdomcentesis total protein
- linear
- the higher the level of protein the more likely it is to have a compromised gut. Values are in gms/dL

23: outcome
- what eventually happened to the horse?
- possible values:
1 = lived
2 = died
3 = was euthanized

24: surgical lesion?
- retrospectively, was the problem (lesion) surgical?
- all cases are either operated upon or autopsied so that this value and the lesion type are always known
- possible values:
1 = Yes
2 = No

25, 26, 27: type of lesion
- first number is site of lesion
1 = gastric
2 = sm intestine
3 = lg colon
4 = lg colon and cecum
5 = cecum
6 = transverse colon
7 = retum/descending colon
8 = uterus
9 = bladder
11 = all intestinal sites
00 = none
- second number is type
1 = simple
2 = strangulation
3 = inflammation
4 = other
- third number is subtype
1 = mechanical
2 = paralytic
0 = n/a
- fourth number is specific code
1 = obturation
2 = intrinsic
3 = extrinsic
4 = adynamic
5 = volvulus/torsion
6 = intussuption
7 = thromboembolic
8 = hernia
9 = lipoma/slenic incarceration
10 = displacement
0 = n/a
28: cp_data
- is pathology data present for this case?
1 = Yes
2 = No
- this variable is of no significance since pathology data is not included or collected for these cases

## Relevant Papers:

N/A

## Papers That Cite This Data Set[1]:


Julie Greensmith. New Frontiers For An Artificial Immune System. Digital Media Systems Laboratory HP Laboratories Bristol. 2003. [View Context].

Richard Nock and Marc Sebban and David Bernard. A SIMPLE LOCALLY ADAPTIVE NEAREST NEIGHBOR RULE WITH APPLICATION TO POLLUTION FORECASTING. International Journal of Pattern Recognition and Artificial Intelligence Vol. 2003. [View Context].

Huan Liu and Hiroshi Motoda and Lei Yu. Feature Selection with Selective Sampling. ICML. 2002. [View Context].

Marc Sebban and Richard Nock and Stéphane Lallich. Stopping Criterion for Boosting-Based Data Reduction Techniques: from Binary to Multiclass Problem. Journal of Machine Learning Research, 3. 2002. [View Context].

Mukund Deshpande and George Karypis. Using conjunction of attribute values for classification. CIKM. 2002. [View Context].

Mark A. Hall. Department of Computer Science Hamilton, NewZealand Correlation-based Feature Selection for Machine Learning. Doctor of Philosophy at The University of Waikato. 1999. [View Context].

Kai Ming Ting and Ian H. Witten. Issues in Stacked Generalization. J. Artif. Intell. Res. (JAIR, 10. 1999. [View Context].

Eibe Frank and Ian H. Witten. Generating Accurate Rule Sets Without Global Optimization. ICML. 1998. [View Context].

Gabor Melli. A Lazy Model-Based Approach to On-Line Classification. University of British Columbia. 1989. [View Context].

H. Altay G uvenir and Aynur Akkus. WEIGHTED K NEAREST NEIGHBOR CLASSIFICATION ON FEATURE PROJECTIONS. Department of Computer Engineering and Information Science Bilkent University. [View Context].

Kai Ming Ting and Ian H. Witten. Stacked Generalization: when does it work. Department of Computer Science University of Waikato. [View Context].

Alexander K. Seewald. Dissertation Towards Understanding Stacking Studies of a General Ensemble Learning Scheme ausgefuhrt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Naturwissenschaften. [View Context].

James J. Liu and James Tin and Yau Kwok. An Extended Genetic Rule Induction Algorithm. Department of Computer Science Wuhan University. [View Context].

## Citation Request:

Please refer to the Machine Learning Repository's citation policy

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info