Annex

Course Project on Advanced Machine Learning

Authors: Arnau Arqué and Daniel Esquina

Description of the problem variables

1. Surgery (categorical). Indicates whether the horse has been treated with surgery or not.
2. Age (categorical). Indicates whether the horse was adult (>=6 months) or young (<6 months).
3. Hospital (string). It indicates the ID of the hospital where you were treated.
4. Rectal temperature (real). Indicates the degrees (celsius) of the horse's rectal temperature.
5. Pulse (real). It codifies the horse's heart rate in beats per minute.
6. Respiratory rate (real). It quantifies the respiratory rate of the horse.
7. Temperature of extremities (categorical). It is a subjective indication of the horse's peripheral circulation coded in Normal, Warm, Cool and Cold.
8. Peripheral pulse (categorical). It is an annotation by healthcare professionals regarding the quality of the horse's blood circulation coded as Normal, Increased, Reduced or Absent.
9. Mucous membranes (categorical). Indicative of the mucous membranes of the horse. It is coded between NormalPink, BrightPink, PalePink, PaleCyanotic, BrightRed and DarkCyanotic.
10. Capillary refill time (categorical). It is a clinical analysis that indicates the circulation capacity of the horse's blood. Encoded in Less3Secs and MoreEqual3Secs.
11. Pain (categorical). It is a subjective indicator of the horse's pain coded in Alert/NoPain, Depressed, Intermittent/MildPain, Intermittent/SeverePain and Continuous.
12. Peristalsis (categorical). Indicative of the intestinal activity of the horse. A low level of this indicator may imply some kind of disease in the horse.
13. Abdominal distension (categorical). It indicates the horse's abdominal pain. An animal with abdominal distension usually has a lot of pain and low intestinal activity. Values can be None, Slight, Moderate or Severe.
14. Nasogastric tube (categorical). It is indicative of gas coming out of the nasogastric tube. It can take the values None, Slight and Significant.
15. Nasogastric reflux (categorical). Indicates the amount of nasogastric reflux of the animal measured in None, More1Liter and Less1Liter.
16. pH of nasogastric reflux (real). pH of the nasogastric reflux measured in the measurement specified above.
17. Rectal examination (categorical). It indicates the presence of faeces in the horse's rectum using the levels Normal, Increased, Decreased and Absent.

18. Abdomen (categorical). It indicates the state of the abdomen of the animal in question. It is measured in Normal, Other, FirmFecesLargeIntestine, DistendedSmallIntestine and DistendedLargeIntestine.

19. Packed cell volume (real). Real value indicating the number of red cells by volume in the blood of the horse.

20. Total protein (real). It is an indicator of the level of dehydration of the horse.

21. Abdominocentesis appearance (categorical). It indicates the appearance of the fluid of the abdominal cavity of the animal. Codified with Clear, Cloudy or Serosanguinous.

22. Abdomcentesis total protein (real). To assess the probability of having a compromised gut.

23. Outcome (categorical). It shows what eventually happened to the horse: Lived, Died or Euthanised.

24. Surgical lesion (categorical, binary). It tells us if the lesion or problem of the horse was surgical with possible values Yes or No.

25-27. Type of the first, second (if any) and third (if any) lesion of the horse. It is codified with $\langle 1 \dots 6 \rangle$ digits indicating $\langle 1, 2 \rangle$ the site, $\langle 3 \rangle$ the type, $\langle 4 \rangle$ the subtype and $\langle 5, 6 \rangle$ the specific code of each of the lesions, respectively. The possible values of each codification are:
    - Site: Gastric, SmGastric, LgColon, LgColonAndCecum, Cecum, TransverseColon, DescendingColon, Uterus, Bladder, AllIntestinalSites, None.
    - Type: Simple, Strangulation, Inflamation, Other.
    - Subtype: Mechanical, Paralytic, None.
    - Code: Obturation, Intrinsic, Extrinsic, Adynamic, Volvulus, Intussuption, Thromboembolic, Hernia, Lipoma, Displacement, None.

28. CP data (categorical, binary). Indicates (Yes or No) whether the pathology data of the animal was present for the case in question. However, the authors pointed out that this variable is of no significance, since the pathology data is not included in this dataset.