**RAUF KEREM YARALI**

**2010.01.05.001**

**HORSE COLİC DATA SET**

About Data Set: A data set about horse diseases. The task is to determine if the lesion of the horse was surgical or not.

**Attribute Information**

1)Surgery:(Nominal data) yes,no

2)Age:(Nominal) adult,young

3)Rectal\_temperature:(Numeric)

4)Pulse:(Numeric)

5)Respiratory\_rate:(Numeric)

6)Temp\_extremities:(Nominal) cool,normal,warm

7)peripheral\_pulse:(Nominal) 1=normal,2=increased,3=reduced,4=absent

8)Mucous\_membranes : (Nominal) normal pink,pale cyanotic,dark cyanotic,pale pink,normal pink,bright pink,

9)Capillary\_refill\_time:(Nominal) 1 <3 , 2 >=3 , 3 meaning unknown

10)Pain:(Nominal) continous severe pain,intermittent mild pain,depressed,alert ,no pain

11)Peritalsis:(Nominal) absent,hypomotile,normal

12)Abdominal\_distension:(Nominal) moderate,severe,slight,none,

13)nasogastric\_tube:(Nominal) 1=none,2=slight,3=significant

14)Nasogastric\_reflux:(Nominal)

15)Nasogastric\_reflux\_PH:(Numeric)

16)Rectal\_examination:(Nominal) decreased,absent,normal,increased

17)Abdomen:(Nominal)distended large intestine,other,normal,firm feces in large intestine,

18)packet\_cell\_volume:(Numeric)

19)total\_protein:(Numeric)

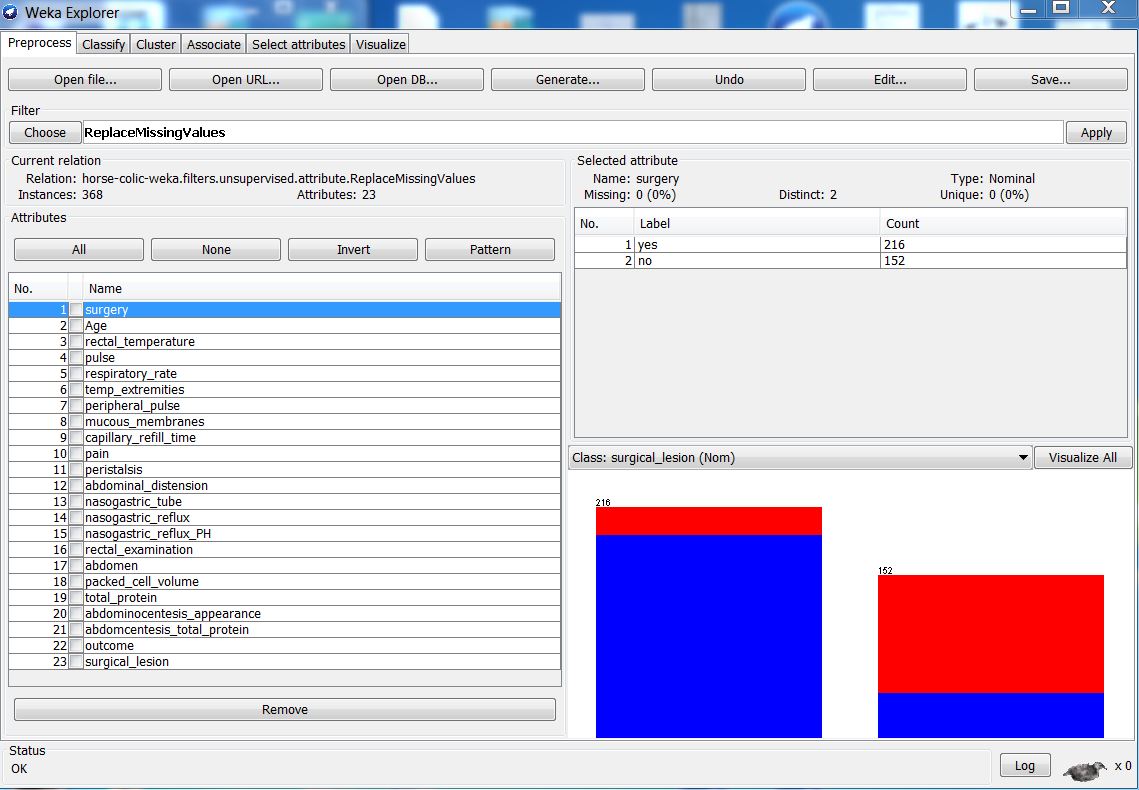
20)Abdominocentesis\_apppearance:(Nominal) cloudy,serosanguinous,clear

21)Abdominoncentesis\_total\_protein:(Numeric)

22)Outcome(Nominal):Died,Euthanized,Lived

23)Surgical\_lesion(Class label):Yes,No

|  |  |
| --- | --- |
| **Data Set Characteristics:** Multivariate | **Number of Instances:** 368 |
| **Attribute Characteristics:** Categorical,Integer | **Number of Attributes:** 22 |
| **Associated Task:** Classification | **Missing Values:** Yes |



Missing Value Recovery : Filter 🡪Choose 🡪 unsupervised 🡪attribute🡪ReplaceMissingValues

Missing values for numeric attributes are replaced with the global mean of each numeric attribute and missing values for nominal attributes are replaced with global mode of each nominal attribute.

**Classification Methods**

**1.Decision Tree(J48)**

*Correctly Classified Instances:* 306 83.1522%

*Incorrectly Classified Instances:* 6216.8478%

*Kappa Statistic: 0.6259*

*Mean Absolute Error: 0.2444*

*Root Mean squared error: 0.3843*

*Relative absolute error: 52.4142%*

*Root Relative squared error: 79.61%*

*Number of leaves:17*

*Size of the tree:24*

*Time taken to build model:0.02 seconds*

*Detail Accuracy By Class*

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

*0.914 0.309 0.835 0.914 0.872 0.787 yes*

*0.691 0.086 0.825 0.691 0.752 0.787 no*

*Weighted avg 0.832 0.227 0.831 0.832 0.828 0.787*

*Confusion Matrix*

*a b 🡪 classified as*

*212 20 |a=yes*

*42 94 |b=no*

*2.Naive Bayes*

*Time taken to build model: 0 seconds*

*Correctly Classified Instances 285 77.4457%*

*Incorrectly Classified Instances 83 22.5543%*

*Kappa Statictic 0.531*

*Mean absolute error 0.2305*

*Root mean squared error 0.4232*

*Relative absolute error 49.4409%*

*Root relative squared error 87.6625%*

*Total number of Instances 368*

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

*0.776 0.228 0.853 0.776 0.813 0.838 yes*

*0.772 0.224 0.669 0.772 0.717 0.838 no*

*Weighted avg. 0.774 0.227 0.785 0.774 0.777 0.838*

*Confusion Matrix*

*a b 🡨 classified as*

*180 52 | a=yes*

*31 105 | b=no*

***RESULTS***

|  |  |  |
| --- | --- | --- |
|  | ***Decision Tree(J48)*** | ***Naive Bayes*** |
| ***Time taken to build model*** | ***0.05 sec*** | ***0.02 sec*** |
| ***Correctly Classified Instances*** | ***306 - 83.152%*** | ***285 – 77.4457%*** |
| ***Incorrectly Classified Instances*** | ***62-16.8478%*** | ***83-22.5543%*** |
| ***Kappa Statistic*** | ***0.625*** | ***0.531*** |
| ***Mean Absolute error*** | ***0.2444*** | ***0.2305*** |
| ***Root mean squared error*** | ***0.3843*** | ***0.4232*** |
| ***Relative absolute error*** | ***52.4142%*** | ***49.4409%*** |
| ***Root relative squared error*** | ***79.6195%*** | ***87.6625%*** |

***According to these results, Decision Tree’s Correctly Classified Instances higher than Naive Bayes.For example,when we look at Decision Tree’s Confision Matrix,we have 212 correctly classified instances for a and we have 94 correctly classified instances for b.When we collect these two instances the result equals to 306.In addition, Decision Tree takes more time than Naive Bayes.***

***Clustering Methods***

***1.Simple kMeans***

***Number of Iterations: 3***

**Within cluster sum of squared errors:2110.0535**

Time taken to build model (full training data) : 0.03 seconds.

Clustered Instances

1. **214 (58%)**
2. **154 (42%)**

**2.Hierarchical Clusterer**

Time taken to build model (full training data): 0.61 seconds

Clustered Instances

**0 367 (100%)**

**1 1 (0%)**

**RESULT**

|  |  |  |
| --- | --- | --- |
|  | **Simple KMeans** | **Hiyerarchical Clusterer** |
| **Time taken to build model** | 0.03 sec | 0.61 sec |

|  |  |  |
| --- | --- | --- |
| **Clustered Instances** | **Simple K Means** | **Hiyerarchical Clusterer** |
| **0** | **214(58%)** | **367(100%)** |
| **1** | **154(42%)** | **1(0%)** |

**SIMPLE kMEANS**

**Classes to cluster:**

**Cluster 0 : no 0 1**

**Cluster1 : yes 98 134 |yes**

**116 20 |no**

**HIERARCHICAL CLUSTER**

**Classes to cluster:**

**0 1 🡨assigned to cluster**

**Cluster 0:yes 231 1|yes**

**Cluster 1:no 136 0|no**