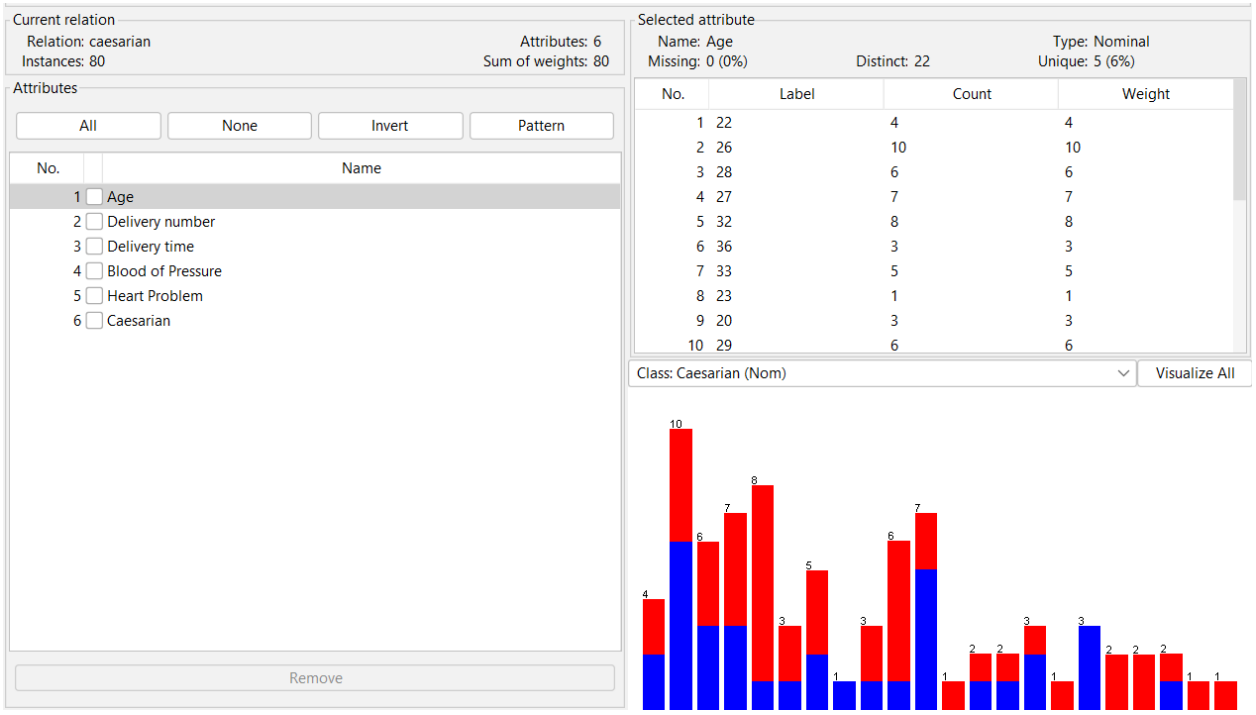
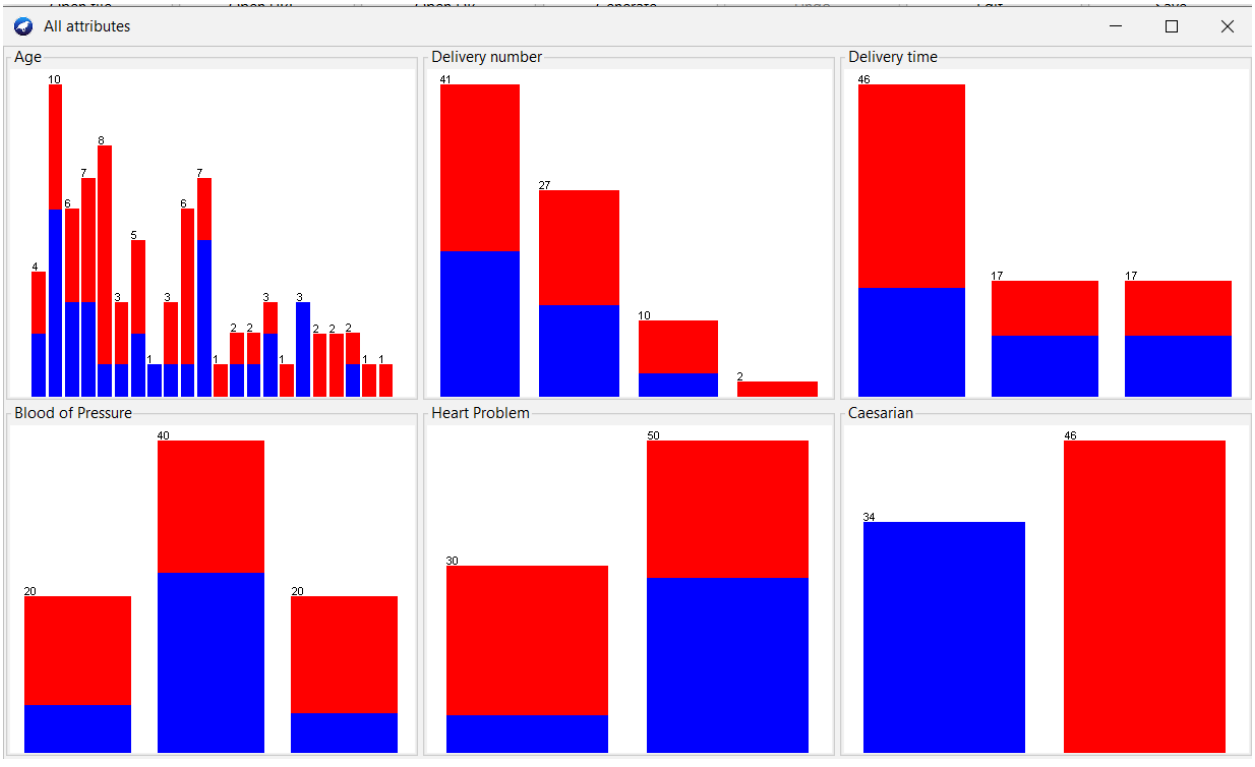
NATURAL LANGUAGE PROCESSING HOMEWORK 3

1. Caesarian Section





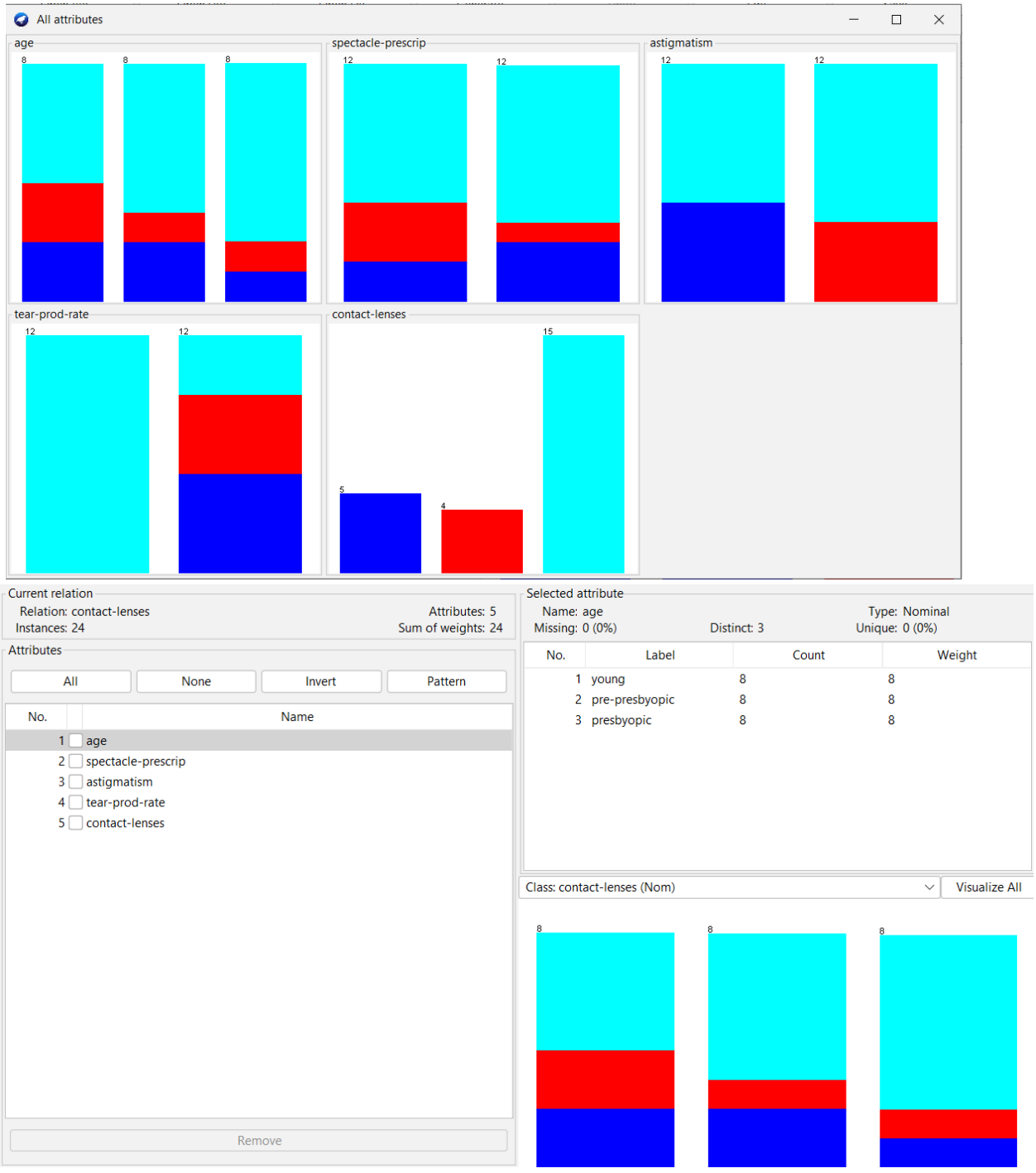
```
              Caesarian
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          46               57.5    %
Kappa statistic                          0
Mean absolute error                      0.4899
Root mean squared error                  0.4952
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000     0,000     ?           0,000   ?          ?        0,439     0,399     0
                 1,000     1,000     0,575       1,000   0,730      ?        0,439     0,549     1
Weighted Avg.    0,575     0,575     ?           0,575   ?          ?        0,439     0,485

=== Confusion Matrix ===

  a  b   <-- classified as
  0 34 |  a = 0
  0 46 |  b = 1
```

Correlation based feature selection ranker method

Attribute selection output

```
Search:         weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:       caesarian
Instances:      80
Attributes:     6
                Age
                Delivery number
                Delivery time
                Blood of Pressure
                Heart Problem
                Caesarian
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 6 Caesarian):
        Correlation Ranking Filter
Ranked attributes:
 0.3526  5 Heart Problem
 0.2393  4 Blood of Pressure
 0.151   3 Delivery time
 0.1001  1 Age
 0.0657  2 Delivery number

Selected attributes: 5,4,3,1,2 : 5
```

2. Contact Lenses



Classifier output

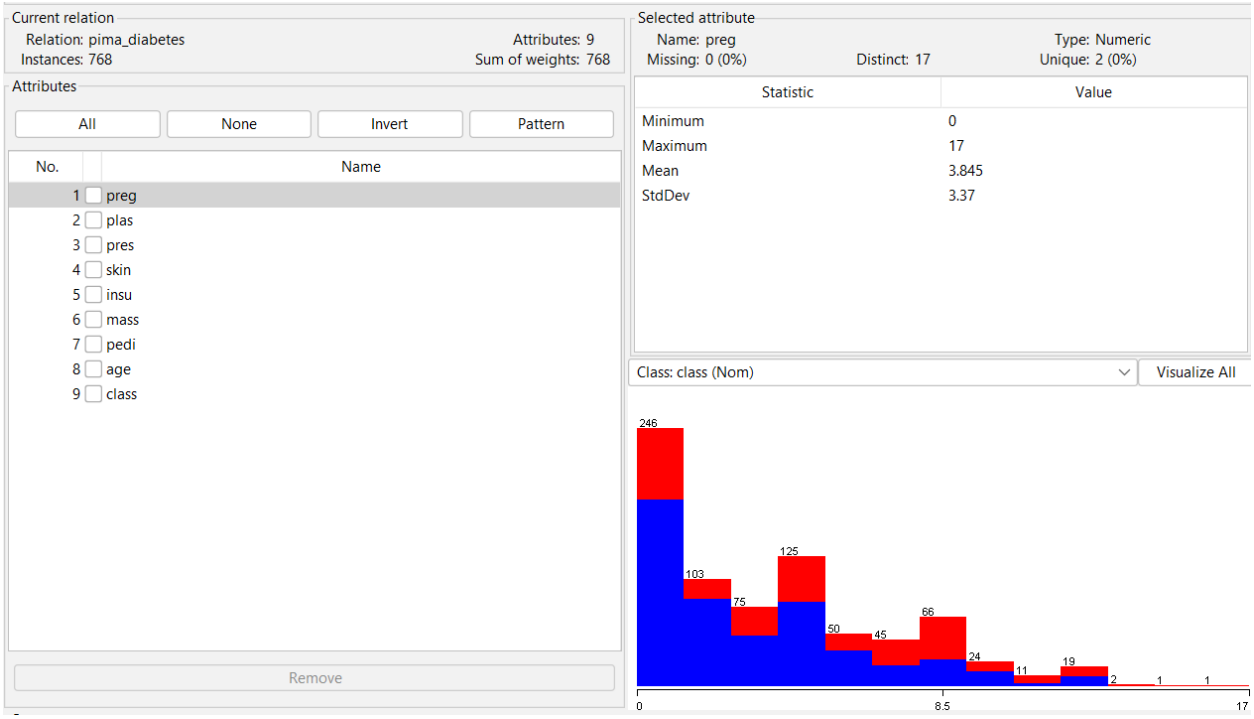
```
=== Classifier model (full training set) ===

ZeroR predicts class value: none

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          15               62.5    %
Kappa statistic                          0
Mean absolute error                      0.3778
Root mean squared error                  0.4367
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000    0,000    ?          0,000    ?          ?        0,216     0,187     soft
                 0,000    0,000    ?          0,000    ?          ?        0,200     0,167     hard
                 1,000    1,000    0,625      1,000    0,769      ?        0,152     0,529     none
Weighted Avg.    0,625    0,625    ?          0,625    ?          ?        0,173     0,398

=== Confusion Matrix ===

  a  b  c   <-- classified as
  0  0  5 |  a = soft
  0  0  4 |  b = hard
  0  0 15 |  c = none
```

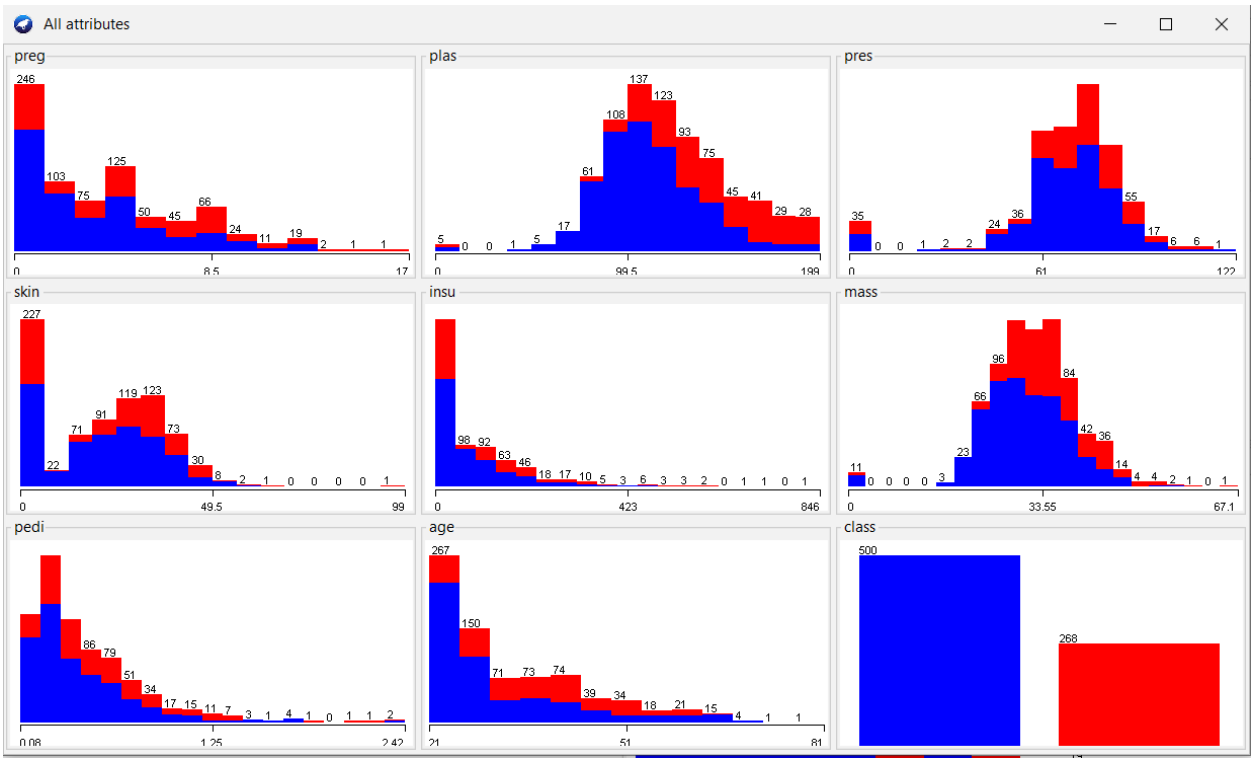
Correlation based feature selection ranker method

Attribute selection output

```
=== Run information ===

Evaluator:    weka.attributeSelection.CorrelationAttributeEval
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     contact-lenses
Instances:    24
Attributes:   5
              age
              spectacle-prescrip
              astigmatism
              tear-prod-rate
              contact-lenses
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 5 contact-lenses):
        Correlation Ranking Filter
Ranked attributes:
 0.666  4 tear-prod-rate
 0.235  3 astigmatism
 0.114  1 age
 0.112  2 spectacle-prescrip

Selected attributes: 4,3,1,2 : 4
```

3. Diabetes


Current relation
Relation: pima_diabetes
Instances: 768
Attributes: 9
Sum of weights: 768
Attributes
All
None
Invert
Pattern
No.
Name
1 preg
2 plas
3 pres
4 skin
5 insu
6 mass
7 pedi
8 age
9 class
Remove
Selected attribute
Name: preg
Missing: 0 (0%)
Distinct: 17
Type: Numeric
Unique: 2 (0%)
Statistic
Value
Minimum 0
Maximum 17
Mean 3.845
StdDev 3.37
Class: class (Nom)
Visualize All
246
103
75
125
50
45
66
24
11
19
2
1
1
0
8.5
17


All attributes
preg
plas
pres
skin
insu
mass
pedi
age
class

Classifier output

```
               class
Test mode:     10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: tested_negative

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         500               65.1042 %
Kappa statistic                          0
Mean absolute error                      0.4545
Root mean squared error                  0.4766
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000     1,000     0,651       1,000   0,789      ?        0,497     0,650     tested_negative
                 0,000     0,000     ?           0,000   ?          ?        0,497     0,348     tested_positive
Weighted Avg.    0,651     0,651     ?           0,651   ?          ?        0,497     0,544

=== Confusion Matrix ===

   a   b   <-- classified as
 500   0 |   a = tested_negative
 268   0 |   b = tested_positive
```

Correlation based feature selection ranker method

Attribute selection output

```
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     pima_diabetes
Instances:    768
Attributes:   9
              preg
              plas
              pres
              skin
              insu
              mass
              pedi
              age
              class
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 9 class):
        Correlation Ranking Filter
Ranked attributes:
 0.4666   2 plas
 0.2927   6 mass
 0.2384   8 age
 0.2219   1 preg
 0.1738   7 pedi
 0.1305   5 insu
 0.0748   4 skin
 0.0651   3 pres

Selected attributes: 2,6,8,1,7,5,4,3 : 8
```