

# DATASETS

## ECOLI3

### General information

| Ecoli (Imbalanced: 3) data set |            |                            |             |
|--------------------------------|------------|----------------------------|-------------|
| Type                           | Imbalanced | Origin                     | Real world  |
| Features                       | 7          | (Real / Integer / Nominal) | (7 / 0 / 0) |
| Instances                      | 336        | IR                         | 8.6         |
| % Positive instances           | 10.42      | % Negative instances       | 89.58       |
| Missing values?                | No         |                            |             |

### Attribute description

| Attribute | Domain               |
|-----------|----------------------|
| Mcg       | [0.0, 0.89]          |
| Gvh       | [0.16, 1.0]          |
| Lip       | [0.48, 1.0]          |
| Chg       | [0.5, 1.0]           |
| Aac       | [0.0,88.0]           |
| Alm1      | [0.03, 1.0]          |
| Alm2      | [0.0, 0.99]          |
| Class     | {positive, negative} |

A imbalanced version of the Ecoli data set, where the positive examples belong to class imU and the negative examples belong to the rest.

# DATASETS

## SEGMENT0

### General information

| Image Segmentation (Imbalanced: 0) data set |            |                            |              |
|---|------------|----------------------------|--------------|
| Type  | Imbalanced | Origin                     | Real world   |
| Features                                    | 19         | (Real / Integer / Nominal) | (19 / 0 / 0) |
| Instances                                   | 2308       | IR                         | 6.02         |
| % Positive instances                        | 14.25      | % Negative instances       | 85.75        |
| Missing values?                             | No         |                            |              |

### Attribute description

| Attribute            | Domain                     | Attribute       | Domain                  |
|----------------------|----------------------------|-----------------|-------------------------|
| Region-centroid-col  | [1.0, 254.0]               | Rawred-mean     | [0.0, 137.11111]        |
| Region-centroid-row  | [11.0, 251.0]              | Rawblue-mean    | [0.0, 150.88889]        |
| Region-pixel-count   | [9.0, 10.0]                | Rawgreen-mean   | [0.0, 142.55556]        |
| Short-line-density-5 | [0.0, 0.33333334]          | Exred-mean      | [-49.666668, 9.888889]  |
| Short-line-density-2 | [0.0, 0.22222222]          | Exblue-mean     | [-12.444445, 82.0]      |
| Vedge-mean           | [0.0, 29.222221]           | Exgreen-mean    | [-33.88889, 24.666666]  |
| Vedge-sd             | [0.0, 991.7184]            | Value-mean      | [0.0, 150.88889]        |
| Hedge-mean           | [0.0, 44.722225]           | Saturatoin-mean | [0.0, 1.0]              |
| Hedge-sd             | [-1.5894573E-8, 1386.3292] | Hue-mean        | [-3.0441751, 2.9124804] |
| Intensity-mean       | [0.0, 143.44444]           | Class           | {positive, negative}    |

A imbalanced version of the Image Segmentation data set, where the possitive examples belong to class 1 and the negative examples belong to the rest.

# DATASETS

## VEHICLE2

### General information

| Vehicle Silhouettes (Imbalanced: 2) data set |            |                            |              |
|--|------------|----------------------------|--------------|
| Type   | Imbalanced | Origin                     | Real world   |
| Features                                     | 18         | (Real / Integer / Nominal) | (0 / 18 / 0) |
| Instances                                    | 846        | IR                         | 2.88         |
| % Positive instances                         | 25.77      | % Negative instances       | 74.23        |
| Missing values?                              | No         |                            |              |

### Attribute description

| Attribute               | Domain              | Attribute          | Domain      |
|-------------------------|---------------------|--------------------|-------------|
| Compactness             | [73, 119]           | Length_rectangular | [118, 188]  |
| Circularity             | [33, 59]            | Major_variance     | [130, 320]  |
| Distance_circularity    | [40, 112]           | Minor_variance     | [184, 1018] |
| Radius_ratio            | [104, 333]          | Gyration_radius    | [109, 268]  |
| Praxis_aspect_ratio     | [47, 138]           | Major_skewness     | [59, 135]   |
| Max_length_aspect_ratio | [2, 55]             | Minor_skewness     | [0, 22]     |
| Scatter_ratio           | [112, 265]          | Minor_kurtosis     | [0, 41]     |
| Elongatedness           | [26, 61]            | Major_kurtosis     | [176, 206]  |
| Praxis_rectangular      | [17, 29]            | Hollows_ratio      | [181, 211]  |
| Class                   | {positive,negative} |                    |             |

A imbalanced version of the Vehicle Silhouettes data set, where the possitive examples belong to class 2 (Bus) and the negative examples belong to the rest.

# DATASETS

## YEAST3

### General information

| Yeast (Imbalanced: 3) data set |            |                            |             |
|--------------------------------|------------|----------------------------|-------------|
| Type                           | Imbalanced | Origin                     | Real world  |
| Features                       | 8          | (Real / Integer / Nominal) | (8 / 0 / 0) |
| Instances                      | 1484       | IR                         | 8.1         |
| % Positive instances           | 10.99      | % Negative instances       | 89.01       |
| Missing values?                | No         |                            |             |

### Attribute description

| Attribute | Domain               |
|-----------|----------------------|
| Mcg       | [0.11, 1.0]          |
| Gvh       | [0.13, 1.0]          |
| Alm       | [0.21, 1.0]          |
| Mit       | [0.0, 1.0]           |
| Erl       | [0.5, 1.0]           |
| Pox       | [0.0, 0.83]          |
| Vac       | [0.0, 0.73]          |
| Nuc       | [0.0, 1.0]           |
| Class     | {positive, negative} |

A imbalanced version of the Yeast data set, where the possitive examples belong to class ME3 and the negative examples belong to the rest.

# DATASETS

## GLASS6

### General information

| Glass Identification (Imbalanced: 6) data set |            |                            |             |
|---|------------|----------------------------|-------------|
| Type  | Imbalanced | Origin                     | Real world  |
| Features                                      | 9          | (Real / Integer / Nominal) | (9 / 0 / 0) |
| Instances                                     | 214        | IR                         | 6.38        |
| % Positive instances                          | 13.55      | % Negative instances       | 86.45       |
| Missing values?                               | No         |                            |             |

### Attribute description

| Attribute | Domain               |
|-----------|----------------------|
| RI        | [1.51115, 1.53393]   |
| Na        | [10.73, 17.38]       |
| Mg        | [0.0, 4.49]          |
| Al        | [0.29, 3.5]          |
| Si        | [69.81, 75.41]       |
| K         | [0.0, 6.21]          |
| Ca        | [5.43, 16.19]        |
| Ba        | [0.0, 3.15]          |
| Fe        | [0.0, 0.51]          |
| Class     | {positive, negative} |

A imbalanced version of the Glass data set, where the positive examples belong to class 6 and the negative examples belong to the rest.