ML Applications using Weka

Sentiment Analysis - A classification task

- Sentiment is feelings
  - Attitudes: positive/negative/neutral?
  - Emotions: happy/sad?
  - Opinions: like/dislike?

- Sentiment analysis: the analysis of feelings behind the words.

- Questions might be asked in sentiment analysis:
  - is this product review positive or negative?
  - how are people responding to this advert campaign?
  - how have bloggers' attitudes about presidential election?

"I am happy with this water bottle."

Positive

"This is a bad investment."

Negative

"I am going to walk today."

Neutral

Sentiment Analysis - A classification task

1. Get text data with labels

2. Extract features from the data, i.e. convert the text items to feature vectors (NLP techniques)

3. Train a model using a ML algorithm for classification, e.g. Neural networks, SVM.

4. Test the model and use it for future classification.

Example: a twitter dataset

- 100 positive tweets and 100 negative tweets from Edinburgh twitter corpus.

```
@relation tweets

@attribute tweet_body string
@attribute sentiment {pos,neg}

@data
'anyone feel motivated the fri afternoon prior to a holiday? wanted to get lots done... but i
want jammies and judge judy... \"SIR!\" &lt;3 her ',pos
'seriously, do you have to rub it in maggie!!!!  ',pos
'if i\'m not wrong.. Alt is when image can\'t be displayed.. Tooltip is the \'title\' ',pos
'I don\'t like social karma  much. Would rather skip it, but can\'t afford to piss my friend off
any more ',pos
'I\'d be happy to review the Iomega if EMC send me one!!! ',pos
```

- Input: tweets, output: positive or negative.

- Try in Weka

Read Weka Outputs

- Running information
  - how the experiment is run
- Classifier mode
  - what the model looks like, e.g. structure, trained parameters, etc.
  - training time
- Performance results
  - Performance summary
  - Detailed performance
  - Confusion matrix (for classification)

Confusion Matrix

- Show statistics of how samples from each class are labelled.

		Predicted condition	
Total 8 + 4 = 12		Cancer	Non-cancer
Actual condition	Cancer	6	2
	Non-cancer	1	3

```
=== Confusion Matrix ===

 a   b    <-- classified as
55  45  |   a = pos
39  61  |   b = neg
```

- Based on the confusion matrix, we can further calculate other metrics, such as recall, precision, etc.

More info can be found here: https://en.wikipedia.org/wiki/Confusion_matrix

Dominant colour extraction - A clustering task

- Sometimes people want to extract colour palettes in artworks.
- What is the colour theme in this image?


```
@relation house
@attribute red real
@attribute green real
@attribute blue real

@data
       26       20       45
       28        5       46
       28       12       44
       28       13       46
       31        4       51
       31        5       48
       31        8       44
       31        9       44
       32        5       50
       32        8       47
       32       11       47
       32       12       48
```

Dominant colour extraction - A clustering task

1. Get RGB values of the image as input.

2. Train a model that cluster RGB vectors into several groups, e.g. K-means.

3. Obtain clusters and output the RGB value of each cluster centre.

- Try in Weka