```
1: ////////////////////////////////////////////////////
2: Machine Learning
3: ____________________________________________________
4: maXbox Starter 60_1 - Sentiment Analysis
5:
6: "Love comes unseen; we only see it go."
7:     - Henry Austin Dobson
8:
9: SA is a way to evaluate and elaborate written or spoken language to
   determine if the expression is favorable, unfavorable, or neutral, and to
   what degree. Sometimes also known as "opinion mining", sentiment analysis
   can let you know if there has been a change in public opinion toward any
   aspect of your business or politics till citizen score or street
   credibility.
10:
11: http://www.softwareschule.ch/examples/machinelearning.jpg
12:
13: To analyze the sentiment of some text, do for example an HTTP POST to
14:
15: http://text-processing.com/api/sentiment/
16:
17: with form encoded data containg the text you want to analyze. You'll get
   back a JSON object response with 2 attributes (label & probability):
18:
19: • label: will be either pos if the text is determined to be positive, neg
   if the text is negative, or neutral if the text as content is neither pos
   nor neg.
20: • probability: an object that contains the probability for each label. neg
   and pos will add up to 1, while neutral is standalone. If neutral is
   greater than the prob. 0.5 then the label will be neutral. Otherwise, the
   label will be pos or neg, whichever has the greater probability. The final
   sentiment is determined by looking at the classification probabilities
   below.
21:
22: The full script is available:
23:
24: http://www.softwareschule.ch/examples/sentiment2.txt
25:
26: Here's some examples using curl & code:
27:
28: $ curl -d "text=great" http://text-processing.com/api/sentiment/
29: {
30:         "probability": {
31:                 "neg": 0.39680315784838732,
32:                 "neutral": 0.28207586364297021,
33:                 "pos": 0.60319684215161262
34:         },
35:         "label": "pos"
36: }
37:
38: function GetSentimentStream4(const S_API, pData: string): string;
39: var ts: TStrings;
40: begin
41:    with TIdHTTP.create(self) do begin
42:       try
43:         ts:= TStringList.Create
44:         ts.Add('text='+HTTPEncode(pData));
45:         result:= Post(S_API,ts);
46:       finally
47:         ts.Free;
48:         Free;
49:       end;
```

```
   end
end;

In a more function oriented way from the HttpPostURL API:

Function HttpPostURL(const URL,URLData: string; const
Data:TStream):Boolean;

We use a StringStream object to pass and a const for the API:

Const
  URLSentimentAPI2='http://text-processing.com/api/sentiment/';

function GetSentimentStream6(const S_API, pData: string): string;
var strm: TStringStream;
begin
  strm:= TStringStream.create('');
  sr:='text='+HTTPEncode(pData);
  if HttpPostURL(S_API, sr, strm) then
     result:= strm.dataString;
  strm.free;
end;


So how Sentiment Analysis with Text Classification Works?

The english sentiment uses classifiers trained on both twitter sentiment
as well as movie reviews from the data sets created by Bo Pang and Lillian
Lee using nltk-trainer (also on bitbucket). The dutch sentiment is based
on book reviews.

The results will be more accurate on text that is similar to original
training data. If you get an odd result, it could be the words you have
used are unrecognized. Try entering more words to improve accuracy.

Those are the 2 parameters
• text: Required - the text you want to analyze. It must not exceed 80,000
characters.
• language: The default language is english, but this API also supports
dutch and french.

And the return values:

On success, a 200 OK response will be returned containing a JSON object
that looks like this:

        "probability": {
                "neg": 0.59797768649386562,
                "neutral": 0.74939503025120124,
                "pos": 0.40202231350613421
        },
        "label": "neutral"

or a shorter one:
{"probability": {"neg": 0.61441529174681542, "neutral":
0.88941013025655569, "pos": 0.38558470825318458}, "label": "neutral"}

A 400 Bad Request response will be returned under the following conditions:

    no value for text is provided or the wrong post form encoded data
    text exceeds 80,000 characters
```

```
102: A 503 Throttled response will be returned if you exceed the daily request
     limit.
103:
104: You can find client libraries for java, ruby, pascal, python, php, and
     objective-c. The public API is for non-commercial purposes, and each
     method is throttled to 1000 calls per day per IP.
105:
106: We are going to create a class called TStringStream that inherits some
     methods from TStream which is why it is passed into the definition of
     HttpPostURL(). This snippet can be used to run allmost all HTTP Post API
     commands independent of language or the  command line tool and use it from
     a script. But sentiment analysis is not a once and done effort.
107:
108: Ref:
109:     Building Machine Learning Systems with Python
110:     Second Edition March 2015
111:
112:     DMath Math library for Delphi, FreePascal and Lazarus May 14, 2011
113:
114:     http://www.softwareschule.ch/box.htm
115:     http://fann.sourceforge.net
116:     http://www.softwareschule.ch/examples/sentiment2.txt
117:
118: Doc:
119:     Neural Networks Made Simple: Steffen Nissen
120:     http://fann.sourceforge.net/fann_en.pdf
121:     http://www.softwareschule.ch/examples/datascience.txt
122:     https://maxbox4.wordpress.com
123:     https://www.tensorflow.org/
124:
125: https://sourceforge.net/projects/maxbox/files/Examples/13_General
126: /811_mXpcatest_dmath_datascience.pas/download
127: https://sourceforge.net/projects/maxbox/files/Examples/13_General
128: /809_FANN_XorSample_traindata.pas/download
129: https://stackoverflow.com/questions/13437402/how-to-run-scrapy-from-within-
     a-python-script
130: https://streamhacker.com/2010/06/16/text-classification-sentiment-analysis-
     eliminate-low-information-features/
131:
132: Plots displaying the explained variance over the number of components is
     called a Scree plot. A nice example of combining a Screeplot with a grid
     search to find the best setting for the classification problem can be
     found at
133:
134: http://scikit-
     learn.sourceforge.net/stable/auto_examples/plot_digits_pipe.html.
135:
136: Although, PCA tries to use optimization for retained variance,
     multidimensional scaling (MDS) tries to retain the relative distances as
     much as possible when reducing the dimensions. This is useful when we have
     a high-dimensional dataset and want to get a visual impression.
137:
138: Machine learning is the science of getting computers to act without being
     explicitly programmed. In the past decade, machine learning has given us
     self-driving cars, practical speech recognition, effective web search, and
     a vastly improved understanding of the human genome. Machine learning is
     so pervasive today that you probably use it dozens of times a day without
     knowing it.
139:
140: >>> https://basta.net/speaker/max-kleiner/
```