pancake

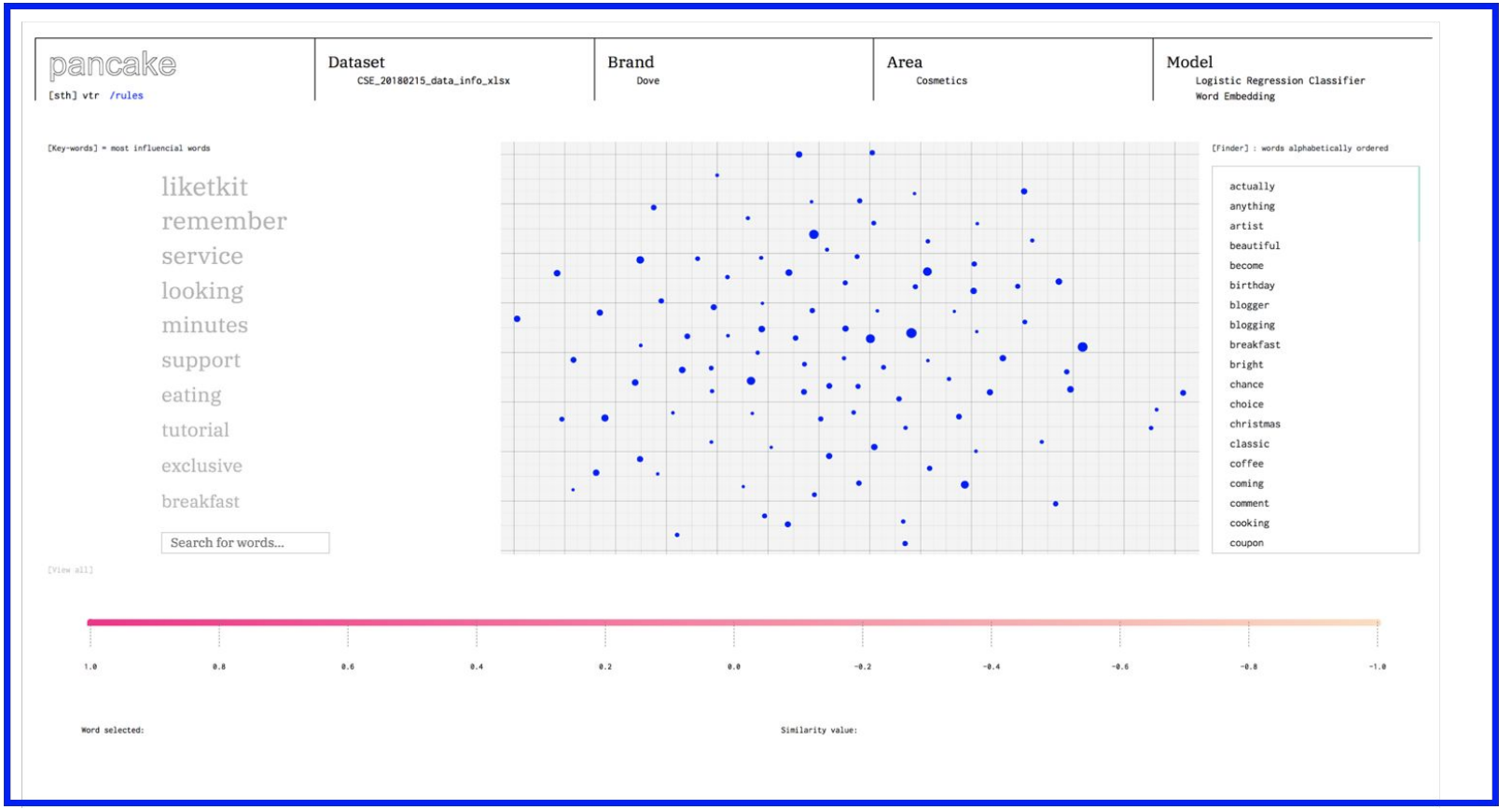
Word-Embedding visual model

Users: Model Developers

Scope: Help the developers to check the model’s result

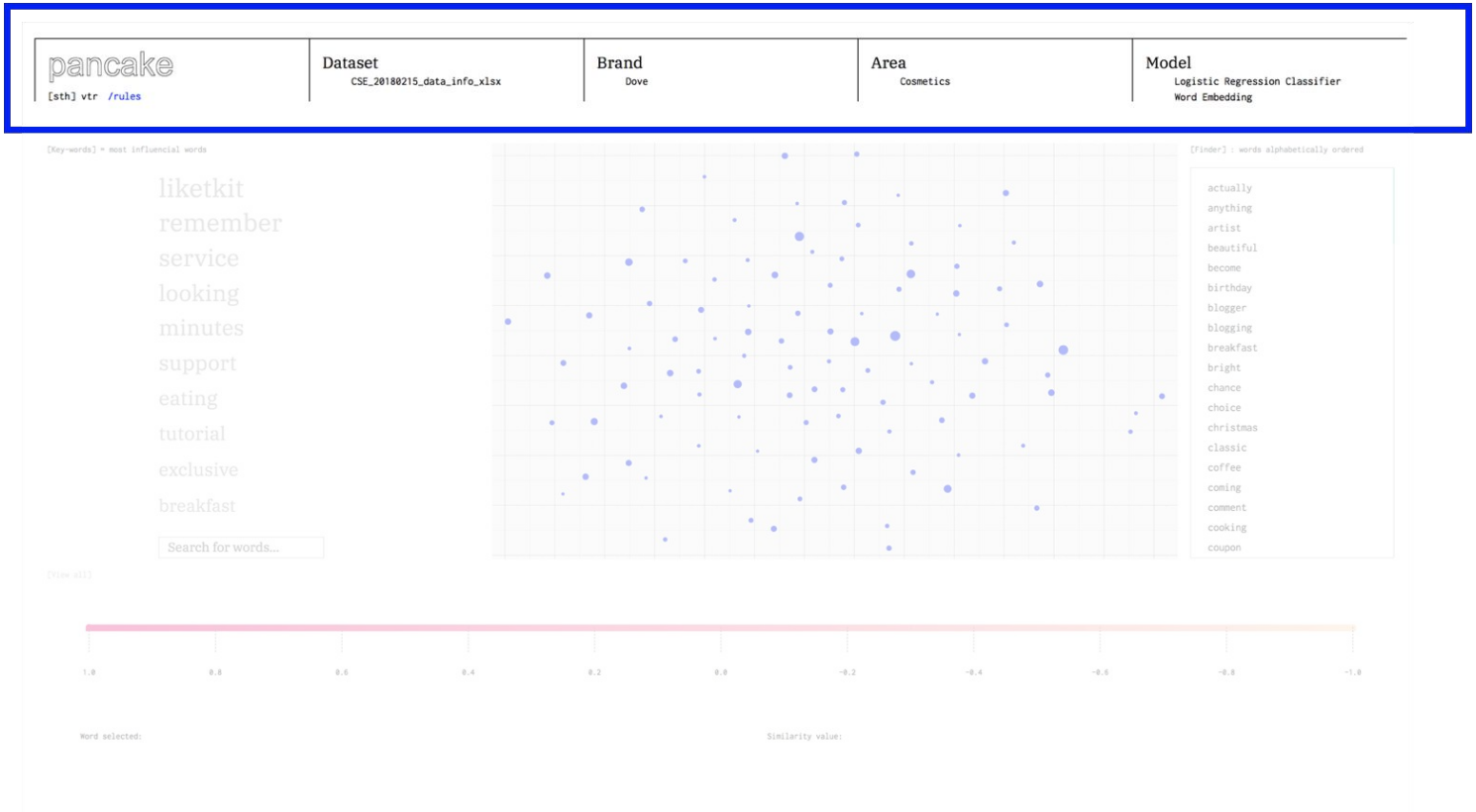
What the tool does: print the words by their similarity

Landing page



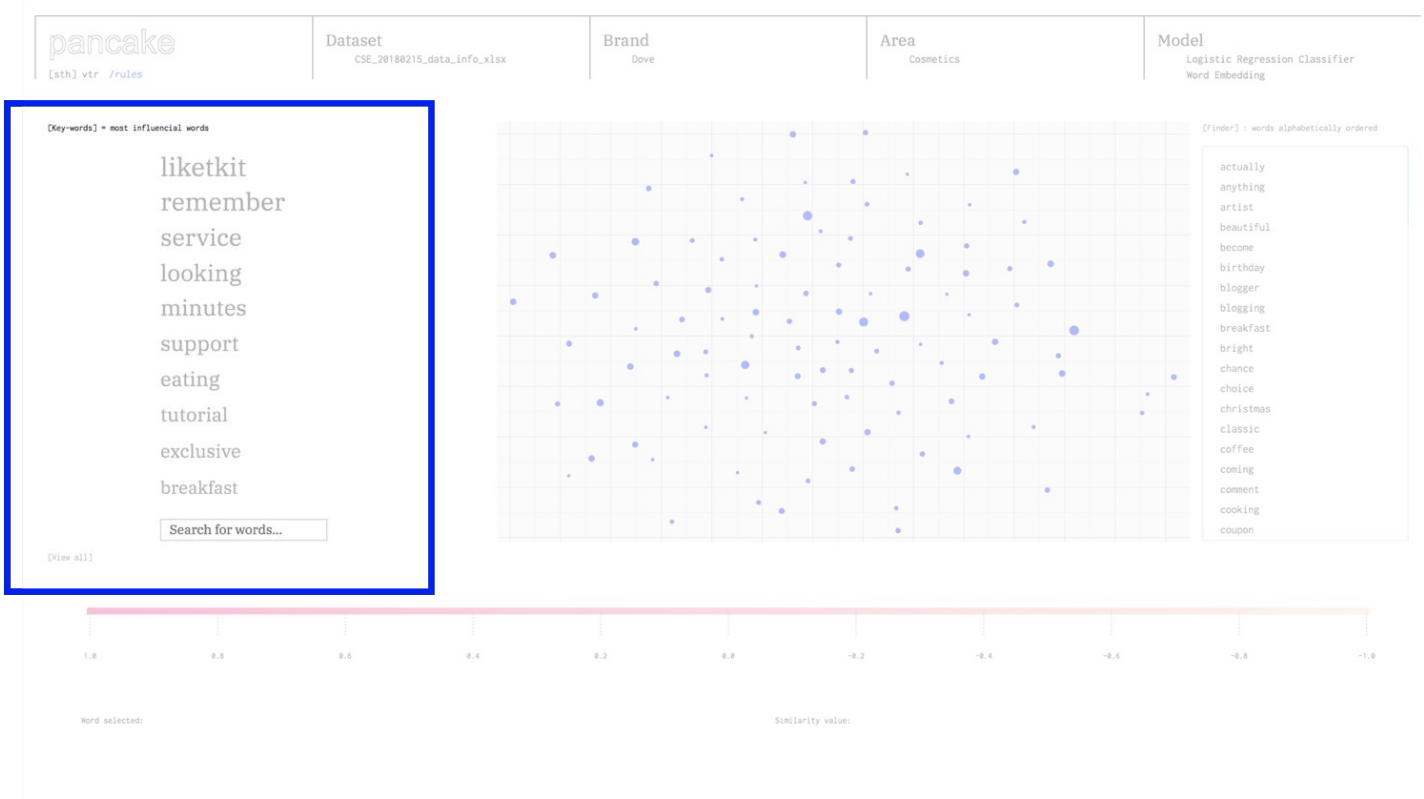
Header

To remind to the developer on which dataset he is is working on and the models that have been applied



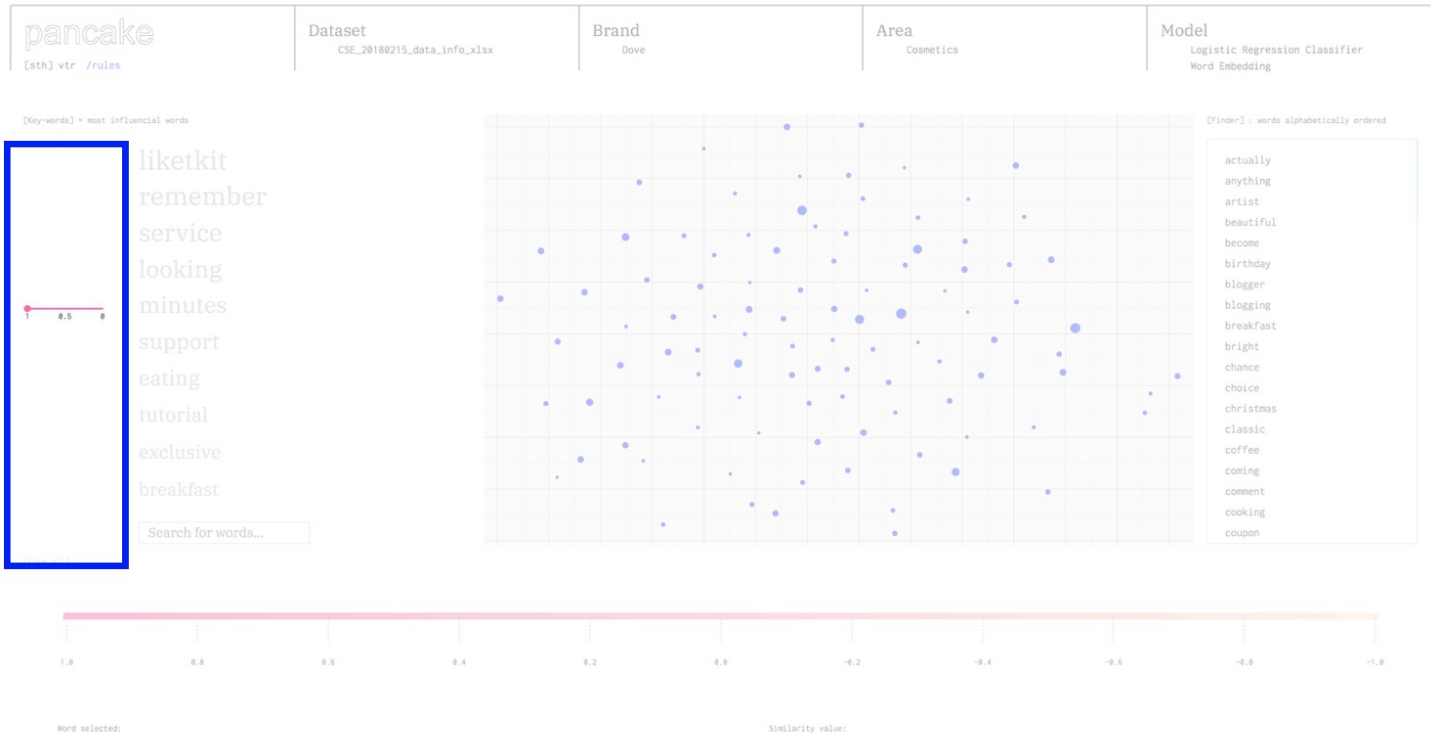
Key-word

List of the most influential words for the text classification task, according to a Logistic Regression Classifier



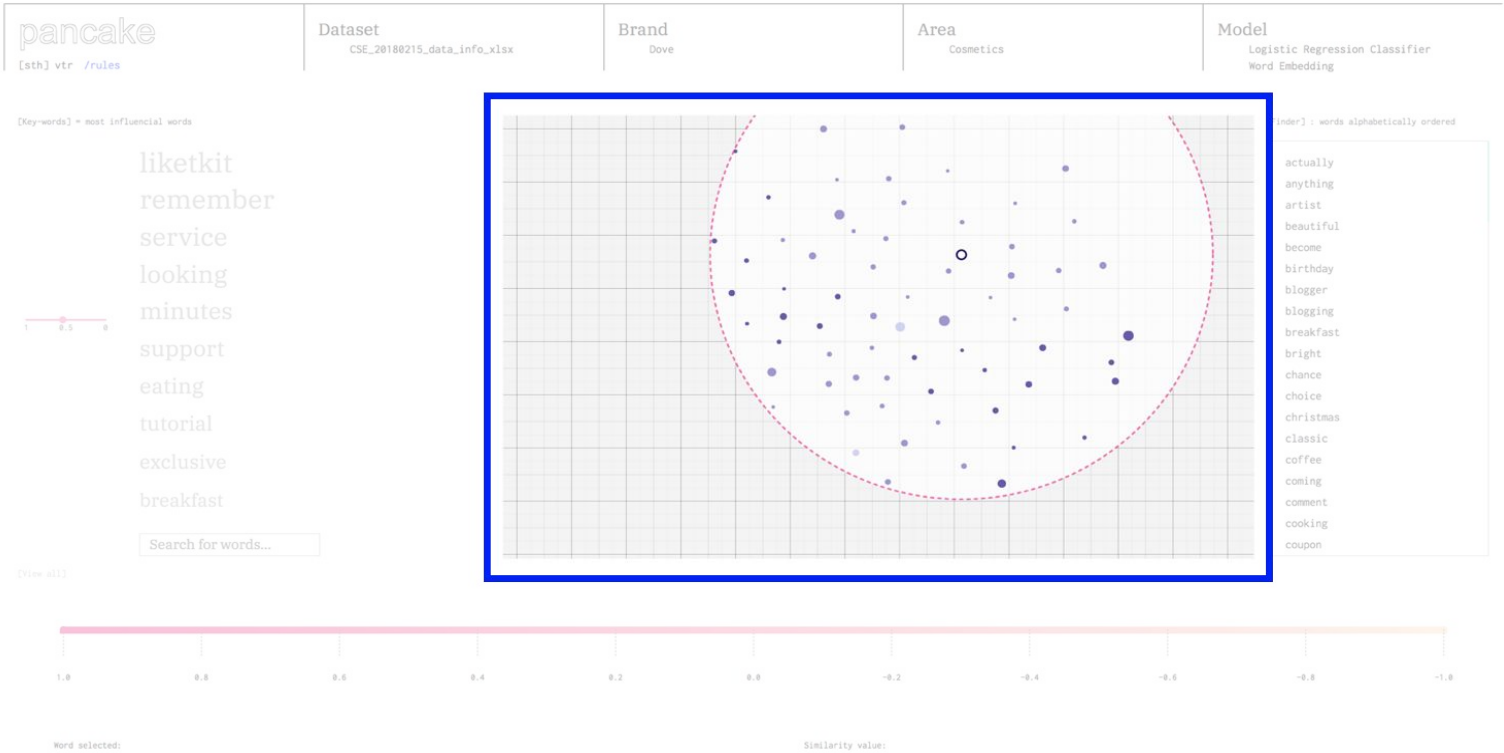
Slide bar

The user can set the similarity range to select the words to analyze



Scatter-plot

Words are positioned in two dimension after having been reduced from 300 to 2 dimensions using TSNE dimensionality reduction



Scatter-plot

How the scatter-plot is built

Dots color

Delta small → Delta large

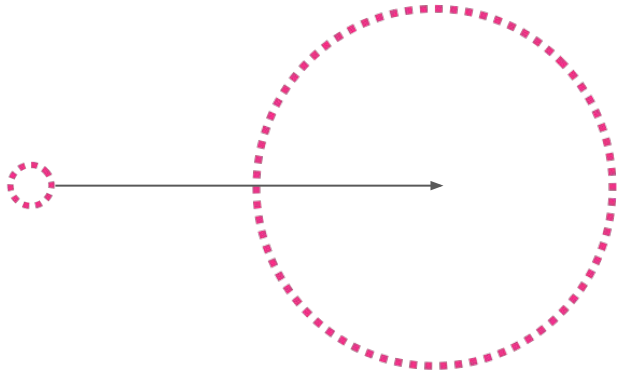
Formula to define dots color:

color_i = delta_i = |cosine_similarity(selected word in 300 dimensions, word_i in 300 dimensions) - cosine_similarity(selected word in 2 dimensions, word_i in 2 dimensions)|

Dots dimension is depending on influence value

High influence → Low influence

Circle selector and slide bar

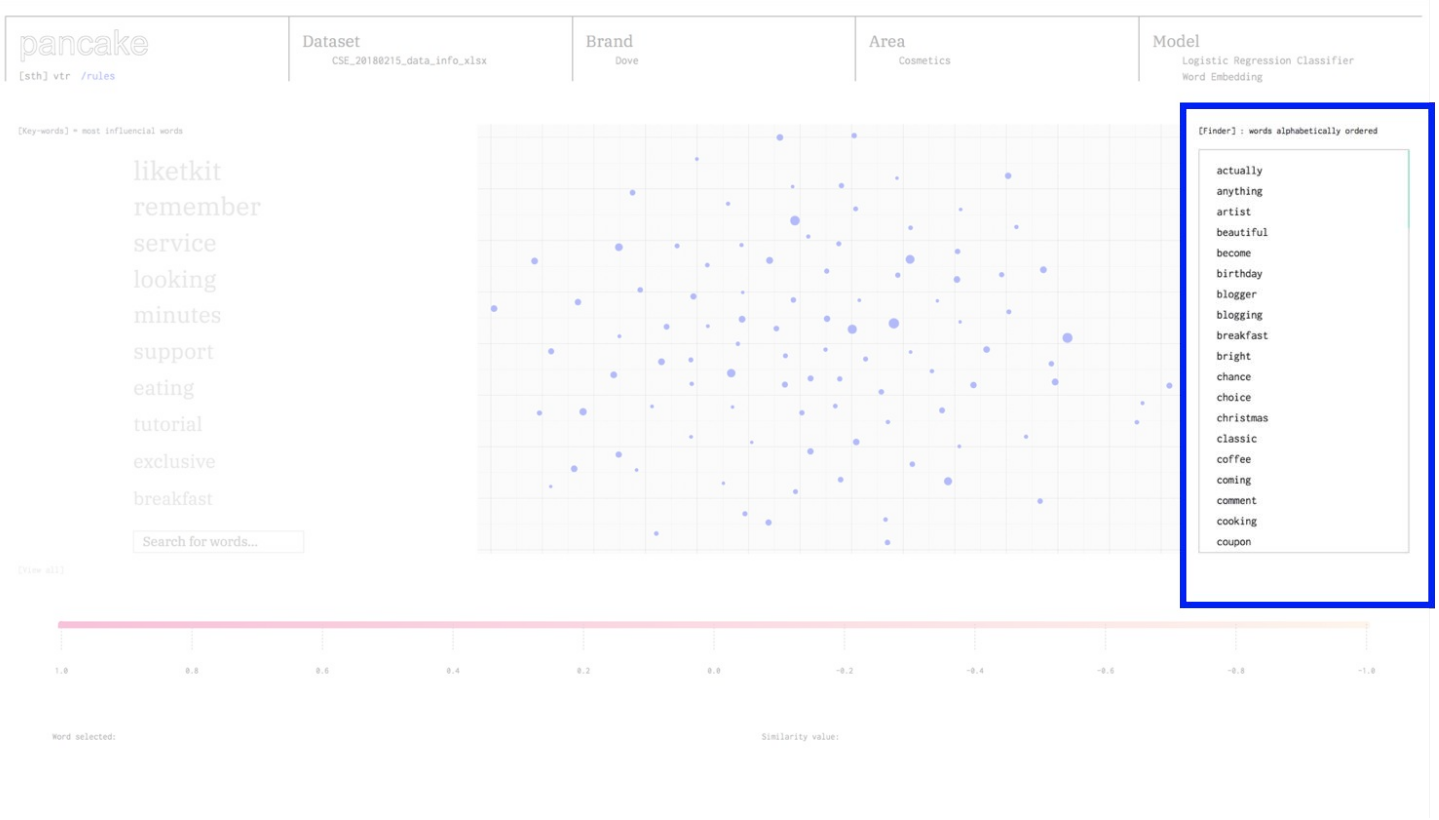


Small selection → Large selection

1 = [corresponding with word selected]

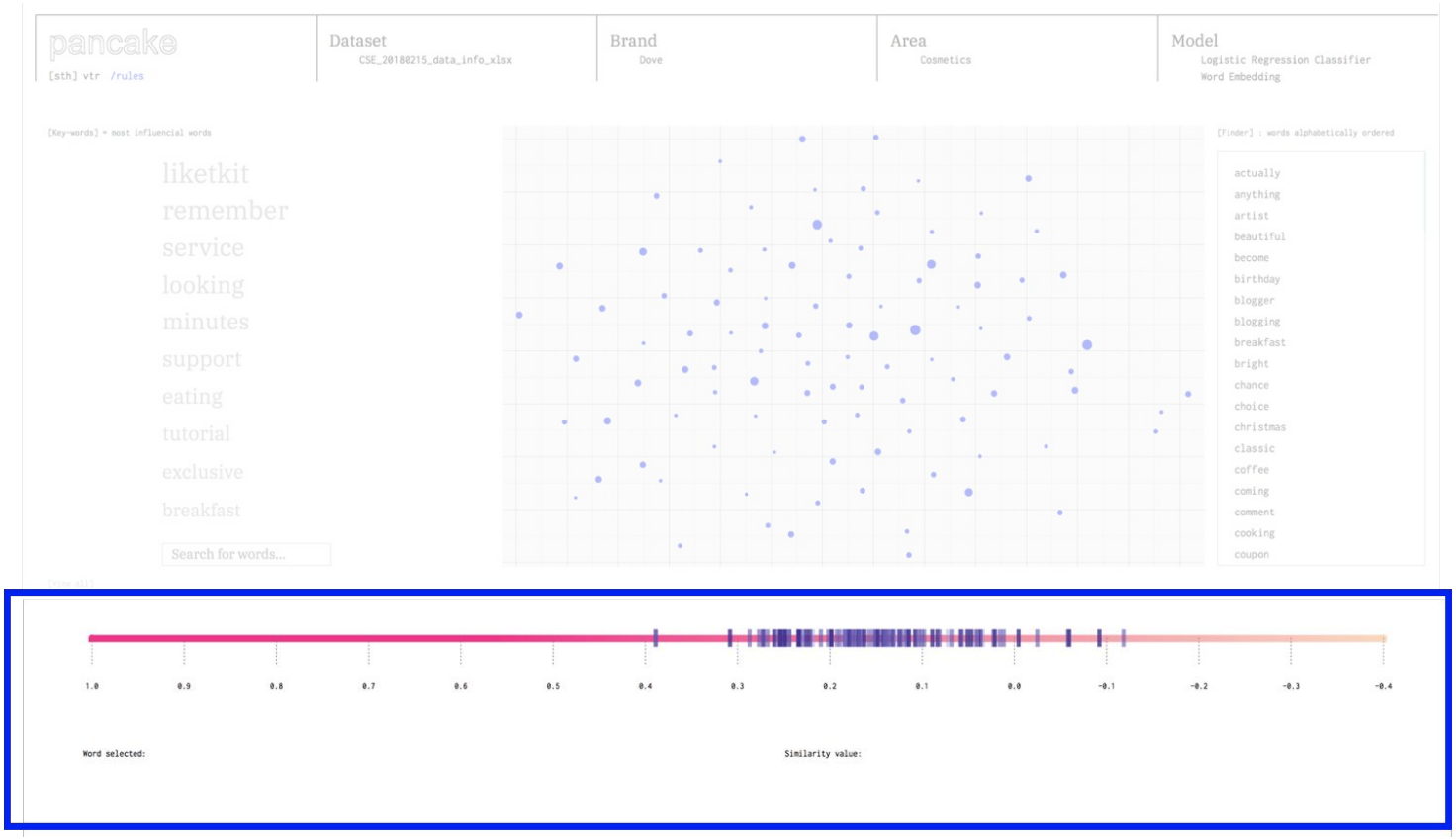
Word-list

List of all the words that have a word embedding representation, ordered alphabetically to ease search



Line-plot

The same words that are plotted on the scatter plot are also plotted in this one dimensional line. The word's distance is plotted according to the cosine similarity in 300 dimensions between the selected words and all the other words with respect to it.



Line-plot

How the line-plot is built

X-axes color

1 = High similarity
[corresponding with word selected]

0 = Low similarity

Rectangles color

Each rectangles is colored like its corresponding dot.
We decided to maintain the same color also if it is not significant in this context, to make easier and faster reading the data equivalence between the two graphs