Explainable Artificial Intelligence Website - User Manual

April 2020

Contents

1 Introduction

This document provides an overview of the Explainable Artificial Intelligence website and instructions on how to browse and interact with the different sections. The website can be accessed at the following address: https://zeno.computing.dundee.ac.uk/2019-projects/matthewfoulis/index.html
This report covers each section of the website in turn. The document can be used to gain an overview of the website as a whole, or the reader can navigate to a specific section for information on a specific topic.

2 Device requirements

The website has been tested to work on the following desktop browsers:

Browser	**Version**
Chrome	81
Firefox	72
Safari	13.1
Edge	44

Table 1: Desktop browsers

and the following mobile browsers:

Browser	**Version**
Safari iOS	13
Firefox iOS	25
Chrome iOS	81
Chrome Android	81

Table 2: Mobile browsers

Note: other browsers may be able to access the website however these are untested. If you have any difficulties accessing the website or would like to report an issue, please see section 8 for contact details.

3 Navigating the website

The website is navigated by scrolling within a webpage, and by using the navigation bar to change between pages. The layout of the navigation bar will change depending on the size of the device you are viewing the website on. In figure 1, the navigation bar is the style seen on a larger device, such as a desktop computer screen. Clicking on the links will navigate to the page chosen, for example, the 'about' page, and clicking on the topic headings will scroll the website to bring that section into view.

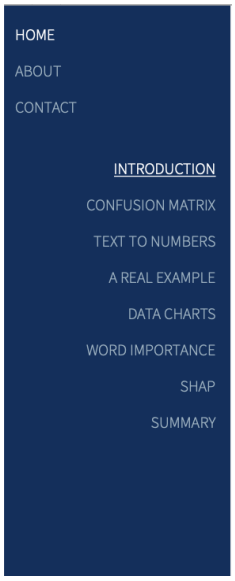

Figure 1: Sidebar as seen when viewing the website on a larger display

In figure 2, the navigation bar is the style seen on a smaller device such as a smartphone.

Figure 2: Sidebar as seen when viewing the website on a smaller display

Clicking on the links will navigate to that page. Note: topic heading links are not presented in the smaller navigation bar style.

4 Home Page

4.0.1 Page Overview

The Home Page is the main page of the website and the location of the majority of the content. The page is divided into separate sections that describe a topic or feature of Machine Learning.

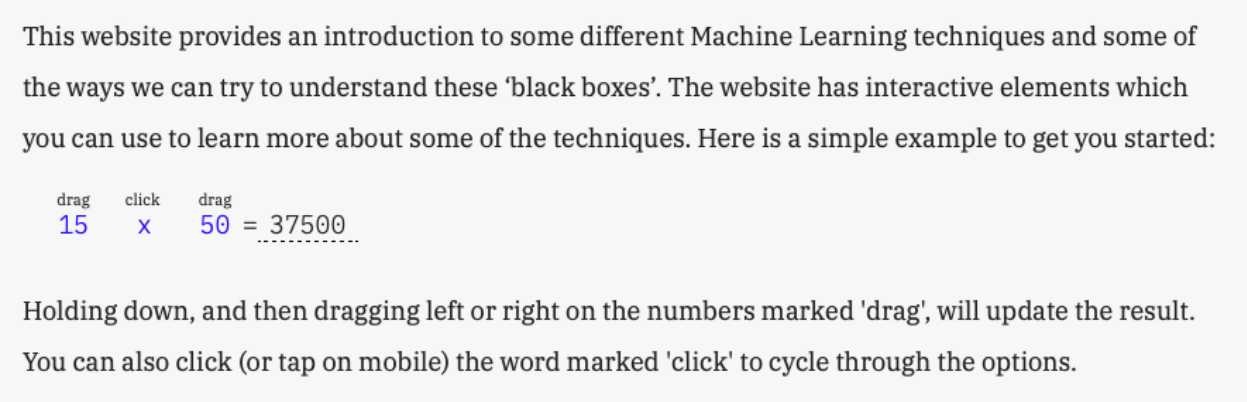

Figure 3: Interactive element within the introduction section of the website

The introduction provides an introduction to the website and the topics to be discussed. Also provided is an example element, see figure 3, which you can interact with to see the effect this has. Initially, you are presented with a multiplication sum and the result of this sum. By holding and dragging you can alter the two numbers. Holding and dragging to the left decreases the number, whilst holding and dragging to the right increases the number. You are also able to click on the mathematical symbol (plus '+', minus '-', and multiply 'x') to change the type of sum. Holding and dragging, or changing the type of sum will update the result accordingly.

4.0.2 Confusion Matrix

In this section, you are introduced to the concept of a confusion matrix. The matrix, as seen in figure 4, can be interacted with to alter the numbers within each of the boxes. By holding and dragging on a number, you can alter the number by dragging to the left and right. Doing so will update the results as seen in figure 5. When you interact with a box within the matrix, the definition for the box will be highlighted in the table to the right. On the results text, clicking on a value will show the definition for that measurement.

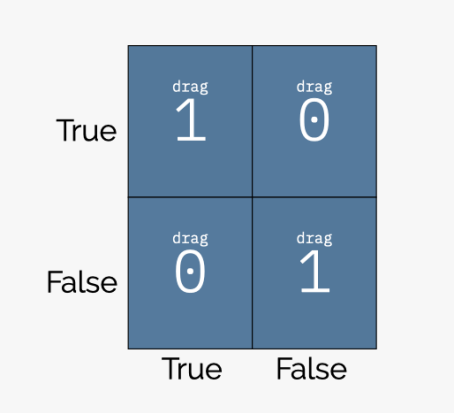

Figure 4: An interactive confusion matrix

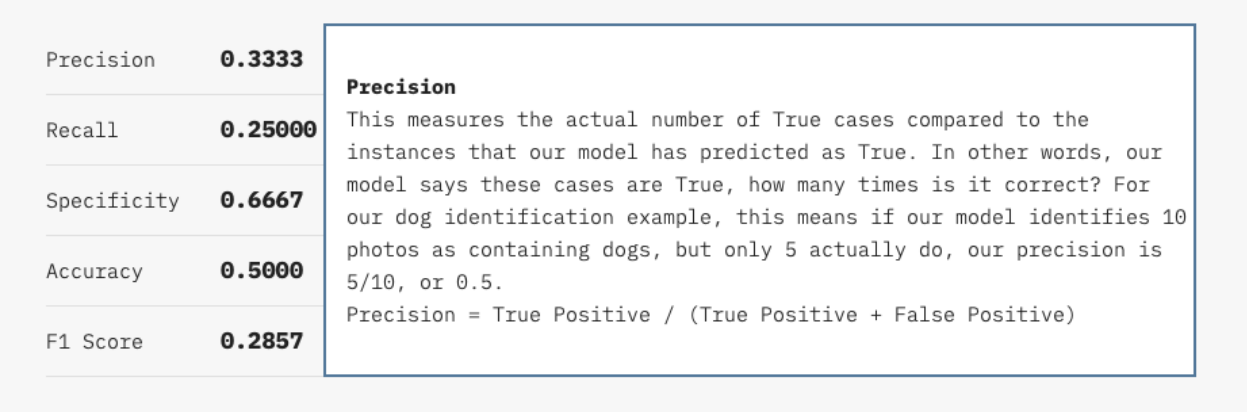

Figure 5: The confusion matrix accuracy measurements and definition for Precision

4.0.3 Text to numbers

In this section, you are introduced to the concept of 'term frequency-inverse document frequency'. In the text, several interactive elements are present. These are blue and underlined with a dotted line, as seen in figure 6.

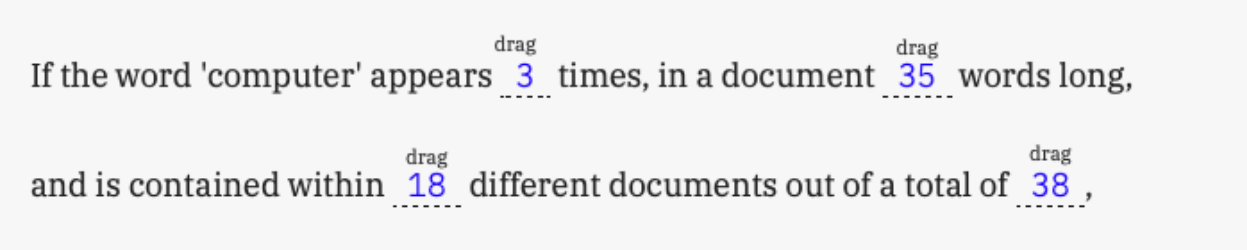

Figure 6: The interactive elements within the term frequency-inverse document frequency section

By holding and dragging these elements, you can change the values and see how this affects the values in the table. An example can be seen in figure 7.

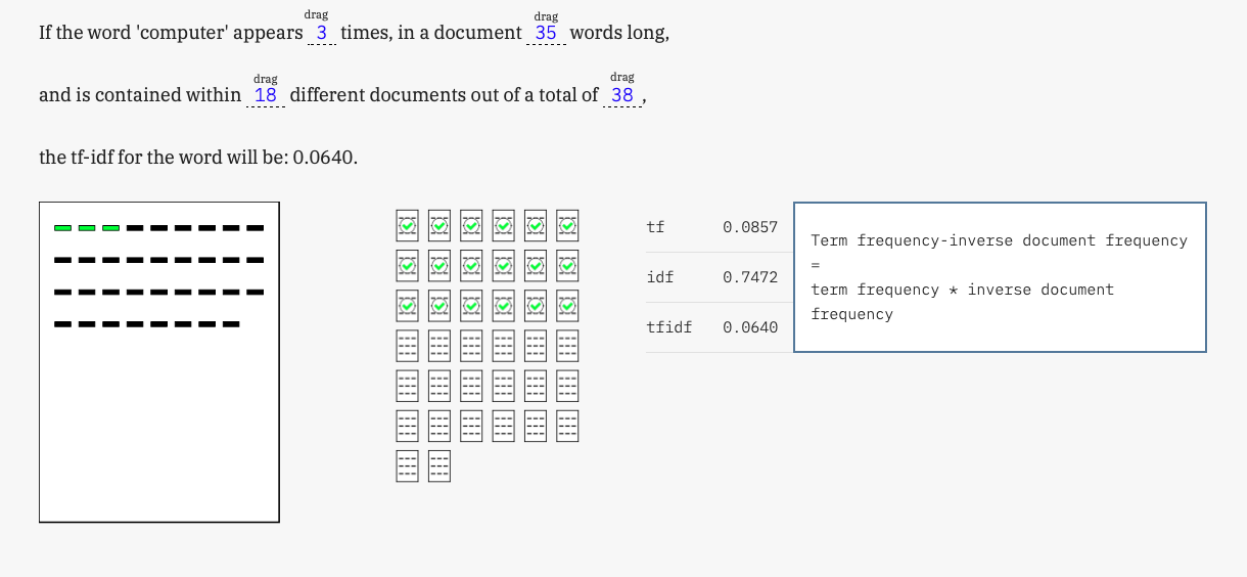

Figure 7: An example from the term frequency-inverse document frequency interactive element

Note: the first and third interactive elements cannot be made greater than the second and fourth elements respectively. The website will prevent you from choosing a number greater than the current second and fourth elements. This is because if there are a total of 20 documents the word ('computer' in this example) could not appear more than 20 times.

4.0.4 Real Data

This section provides a description of previous work carried out in conjunction with the website and provides some context for the later sections. A confusion matrix, as seen in figure 8, showing the results achieved in this work is displayed. Note: as the values displayed are taken from an actual example, the confusion matrix is not interactive.

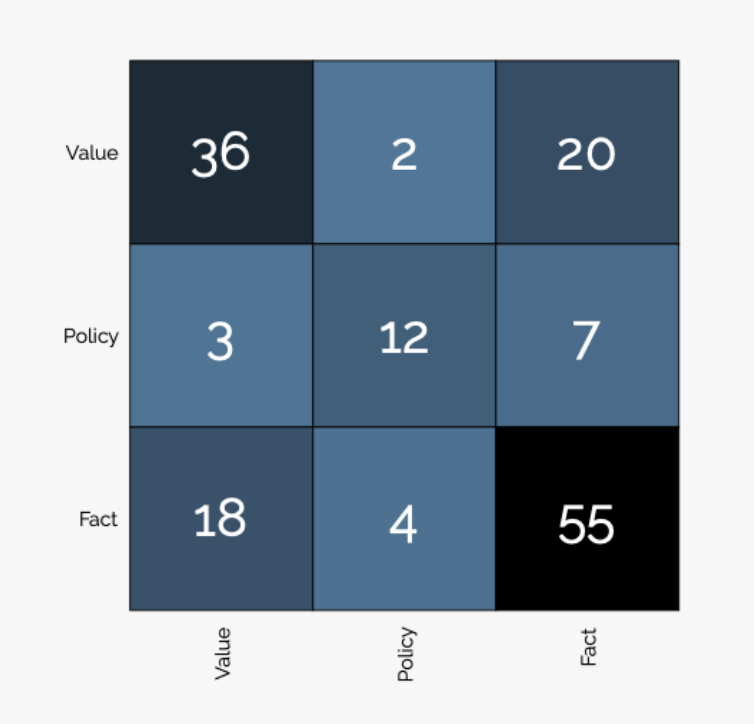

Figure 8: A confusion matrix showing the results in achieved in a proposition classification task.

4.0.5 Scatter Plot

In this section, you can investigate some of the features used by the Machine Learning model when determining its predictions. A scatter plot, as seen in figure 9, displays the values of two features compared with each other on the X and Y-axis. Clicking on the axis label will change the label and the points on the chart will update accordingly. Clicking on the label will update it until all values have been cycled through, at which point it will return to the original value. Hovering over, or tapping on mobile, on a point on the plot will display the values of the X and Y-axis for that point, as well as the text of the proposition that the point represents.

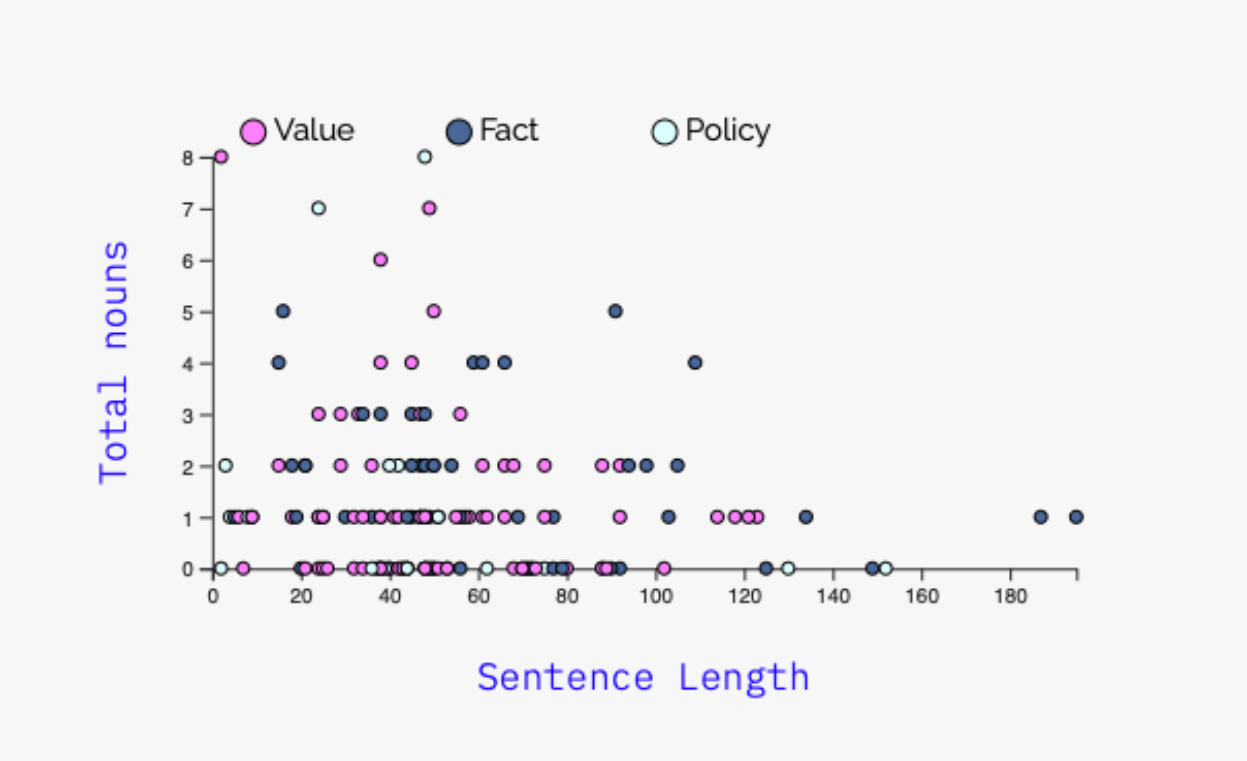

Figure 9: A interactive scatter plot displaying machine learning features

The bar chart, as seen in figure 10, also presents a feature from the data. Clicking on the label on the X-axis will similarly cycle through the features. Hovering over a bar, or tapping on a mobile device, will display the value of the bar (from the X-axis) and the number of occurrences for that value (from the Y-axis).

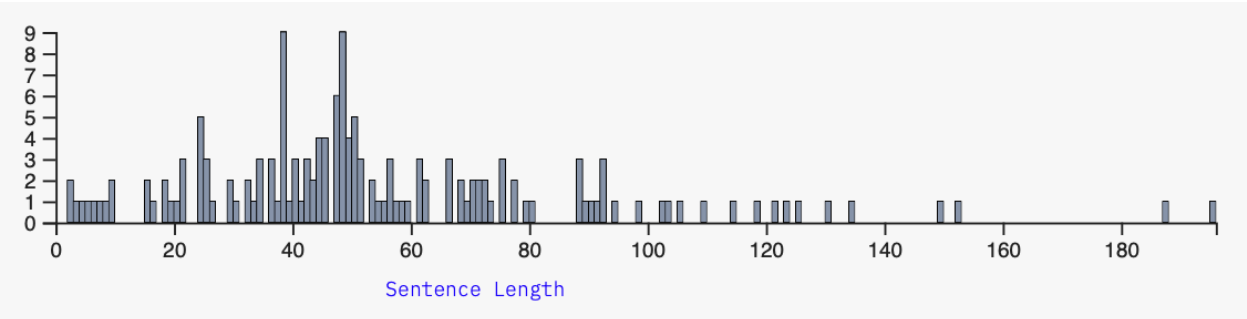

Figure 10: A interactive bar chart displaying machine learning features

Additionally, on the scatter plot, you can click and drag to outline a rectangle, which can be used to select

a set of points on the plot. When a point is highlighted, the corresponding value on the bar chart will be highlighted. An example can be seen in figure 11.

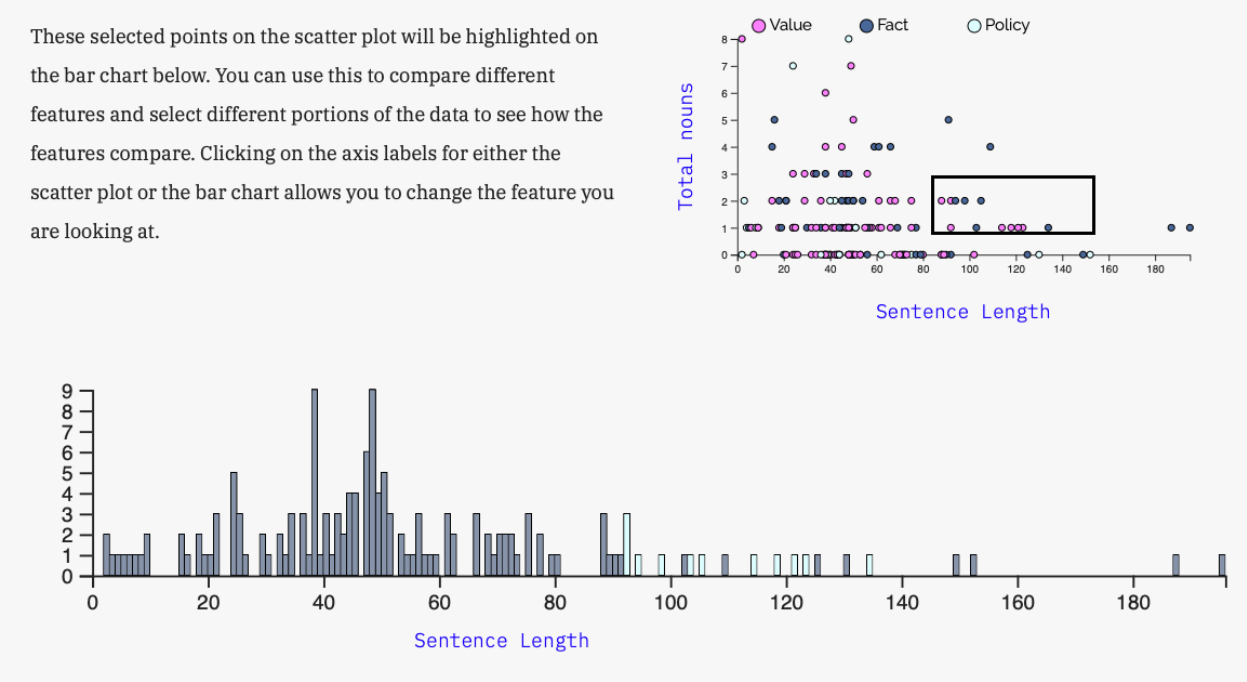

Figure 11: An example of a selection within the scatter plot highlighting elements within the bar chart

4.0.6 Word Importance

This section introduces the concept of Local Interpretable Model-agnostic Explanations (LIME). An introductory text is provided and then you are presented with a set of interactive elements to gain a further understanding of the topic, seen in figure 12.

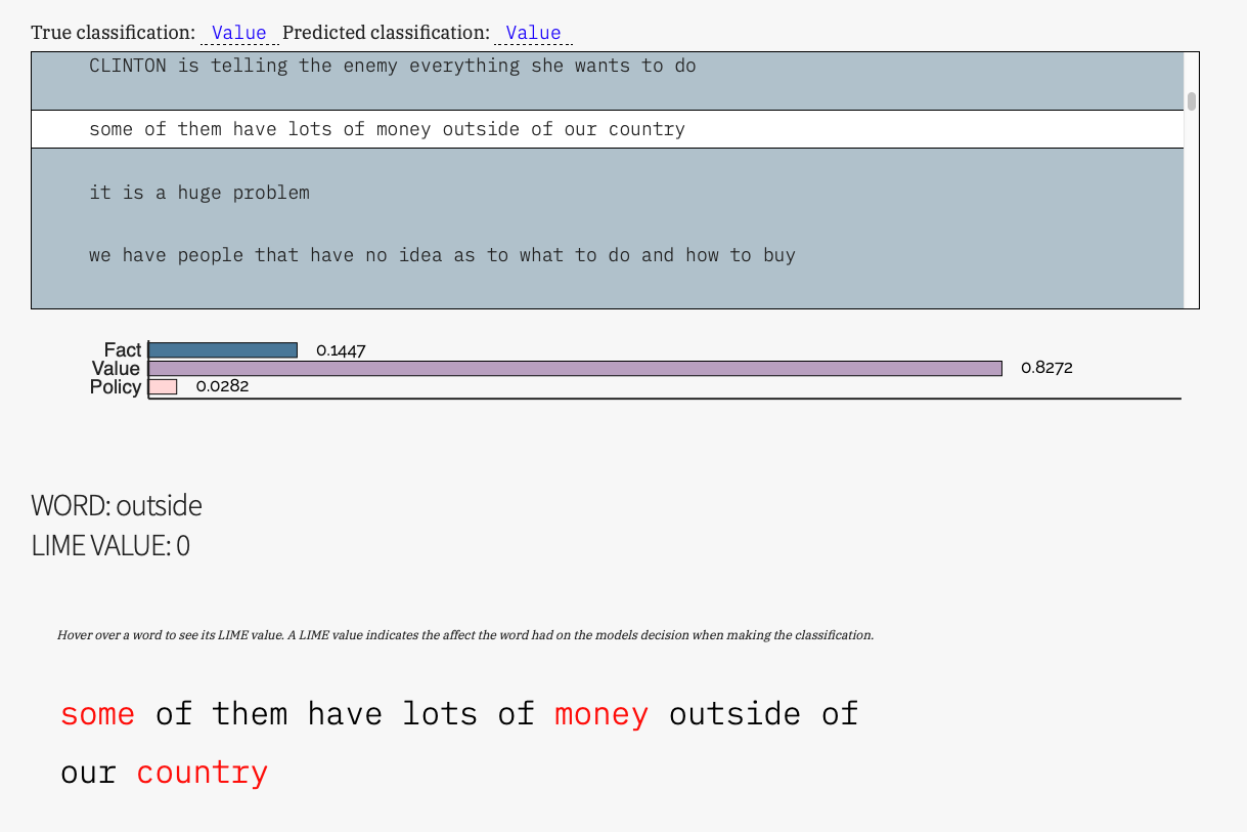

Figure 12: The interactive word importance section of the website

At the top of the element, the two underlined words allow you to alter the inclusion criteria for the list of sentences beneath. Initially, the sentences shown are propositions which have a true class of 'Value', and were correctly classified by the Machine Learning model as 'Value' propositions. By clicking the underlined words, you are able to alter these 'true classification' and 'predicted classification' to show a different set of sentences. For example, changing the words to be 'Fact' and 'Policy', would show propositions which are type Fact, but were classified by the model as Policy.

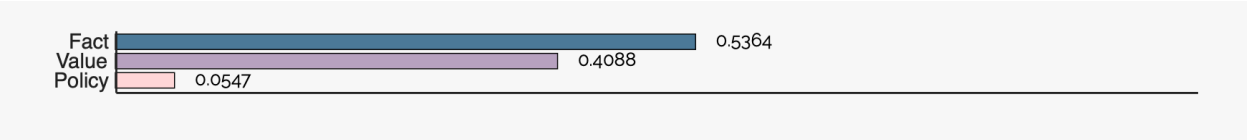

Figure 13: A bar chart displaying the classification likelihood for a chosen proposition

With the propositions displayed, you can click on them to be provided with additional information. When a proposition is selected, the bar chart below will be updated to show the classification probability. For example, the bar chart in figure 13 shows us that the proposition was predicted as type 'Fact' with 0.5364 percent likelihood.
The next element, as seen in figure 14, updates when a new proposition is selected. Hovering over the words, or tapping if using a mobile device, displays the LIME value - the importance the word had to the

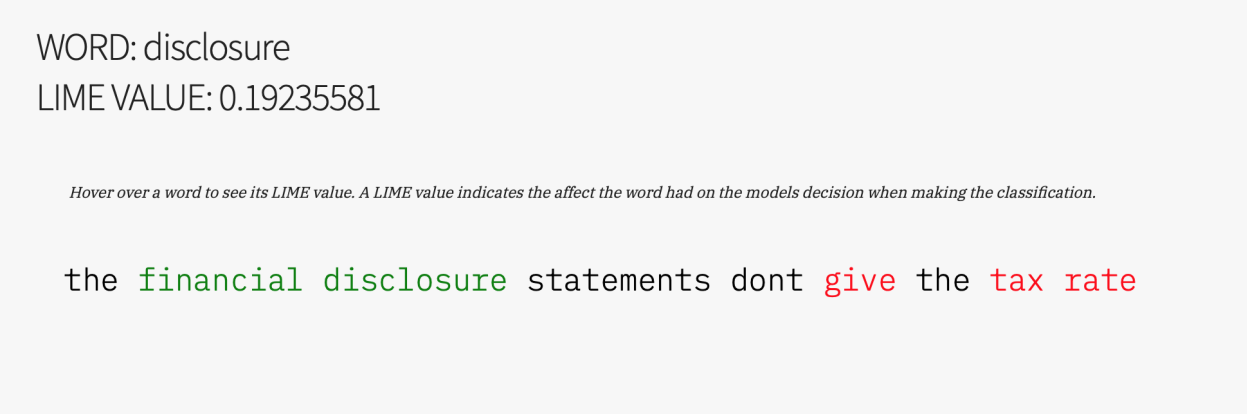

Figure 14: An example of the word importance interactive element

classification - for that individual word. Red words had a negative effect on the classification, and the green words had a positive effect on the classification.

4.0.7 SHAP

This section introduces the concept of Shapley values and how they can be used in an explainable artificial context. An introduction is provided and then a scatter plot displays information achieved using the SHAP framework. It shows the top ten features, in order of importance from top to bottom on the Y-axis; the affect the points had on the predictions, positive or negative, on the X-axis; and the relative value the point has represented by its colour. For example, the point circled in figure 15 is a 'length' feature which is used to measure the length of the sentence. It is positioned at 1.2 on the X-axis so we know it had a positive effect on the prediction (meaning it increased the likelihood of the prediction), and it is red in colour so we know that out of all the sentence lengths, it had a high relative value.

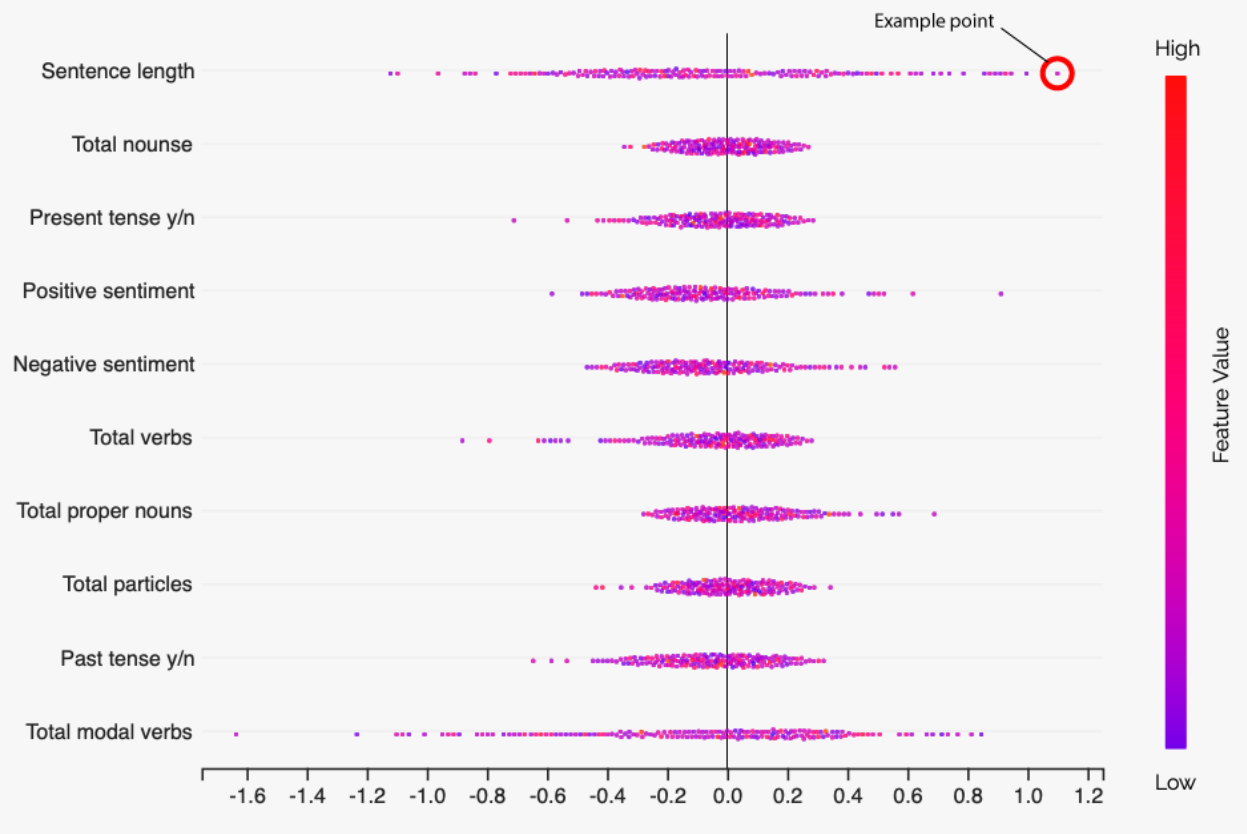

Figure 15: A highlighted point within the SHAP scatter plot

4.0.8 Summary and more information

This section provides additional sources where more information about the topics can be found. Clicking on the hyperlinks, as seen in figure 16, will take you to the chosen website.

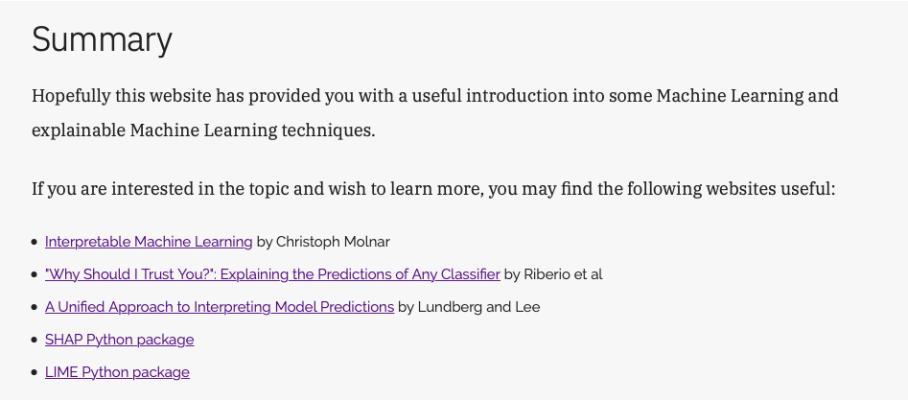

Figure 16: The summary and more information section of the website

5 About Page

5.0.1 Page Overview

The About Page provides a description of the website including why it was created and what the aim of the website is.

6 Contact Page

6.0.1 Page overview

The Contact Page provides contact details for the developer of the website. These may be used to contact the developer regarding a query, to report a technical problem with the website, to request further information, or regarding any other relevant issue.

6.0.2 Using the contact details

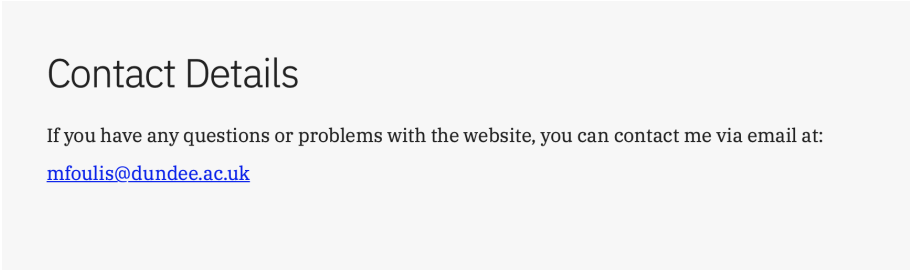

Figure 17: The contact details section of the website

To send an email, the email address, as seen in figure 17, can be clicked to open your default email client with the address pre-filled in the 'to' field. Alternatively, you can right-click, or long press if using a mobile device, the email, and select 'Copy Email Address'. This can then be pasted into the 'to' field within your email client.

7 Troubleshooting

The website does not display properly/certain images or elements do not appear:
For the website to work correctly, Javascript must be enabled. If Javascript is not enabled, certain aspects of the website, such as the interactive elements, will not appear and/or not work correctly.

Javascript is enabled but the website still does not display correctly:
Firstly, please check section 2. If your browser is not listed the website has not been tested for it. Please try another browser or see section 8 for the contact details.

8 Further help

If you require further assistance, please use the following contact details:
email: mfoulis@dundee.ac.uk