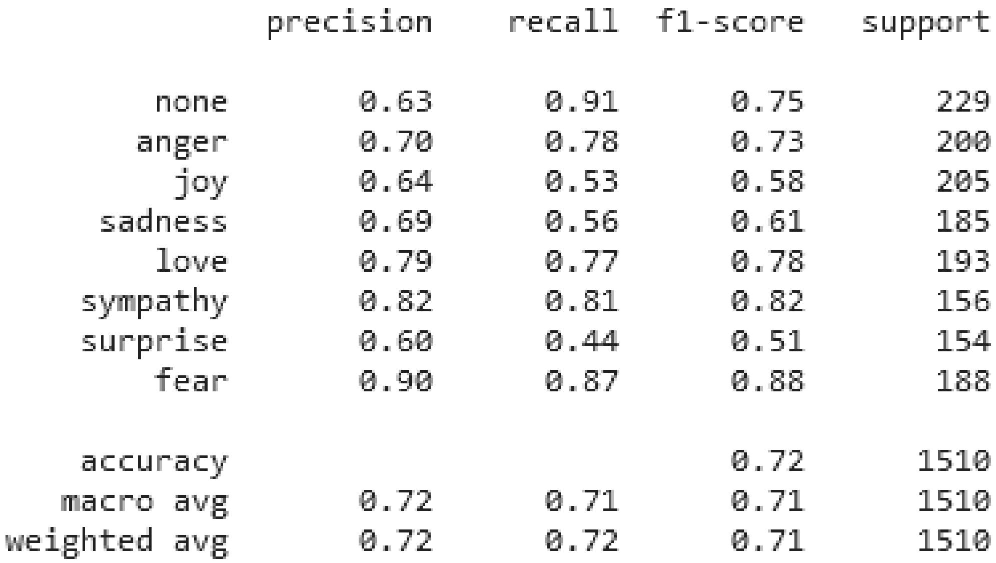
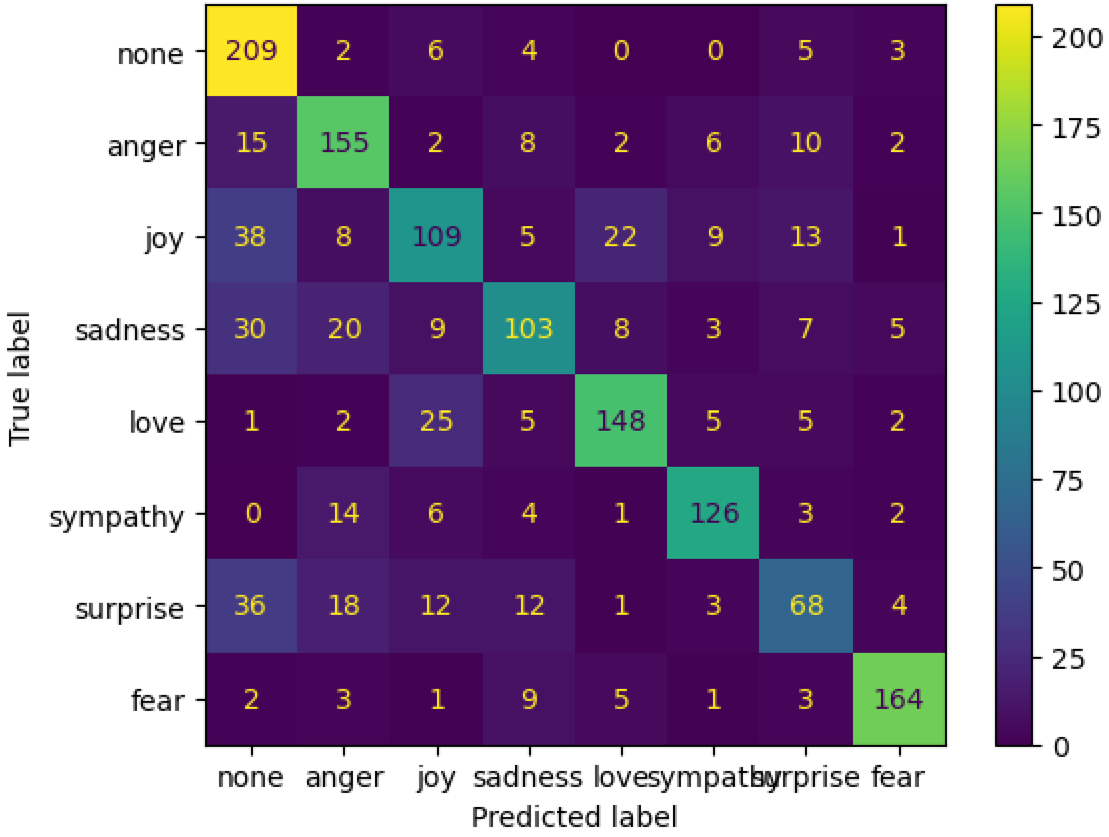
We will be using the light stemming model from now.

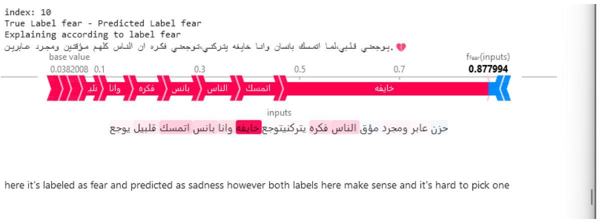


```
              precision    recall  f1-score   support

        none       0.63      0.91      0.75       229
       anger       0.70      0.78      0.73       200
         joy       0.64      0.53      0.58       205
     sadness       0.69      0.56      0.61       185
        love       0.79      0.77      0.78       193
    sympathy       0.82      0.81      0.82       156
    surprise       0.60      0.44      0.51       154
        fear       0.90      0.87      0.88       188

    accuracy                           0.72      1510
   macro avg       0.72      0.71      0.71      1510
weighted avg       0.72      0.72      0.71      1510
```



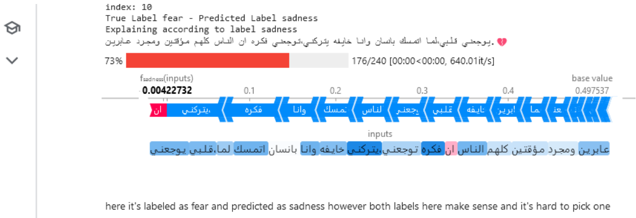
	none	anger	joy	sadness	love	sympathy	surprise	fear
none	209	2	6	4	0	0	5	3
anger	15	155	2	8	2	6	10	2
joy	38	8	109	5	22	9	13	1
sadness	30	20	9	103	8	3	7	5
love	1	2	25	5	148	5	5	2
sympathy	0	14	6	4	1	126	3	2
surprise	36	18	12	12	1	3	68	4
fear	2	3	1	9	5	1	3	164

Light Stemming vs Raw Data:



```
index: 10
True Label fear - Predicted Label fear
Explaining according to label fear
يوجعني قلبي،لما اتمسك بانسان وانا خايفه يتركني،توجعني فكره ان الناس كلهم مؤقتين ومجرد عابرين.💔
```

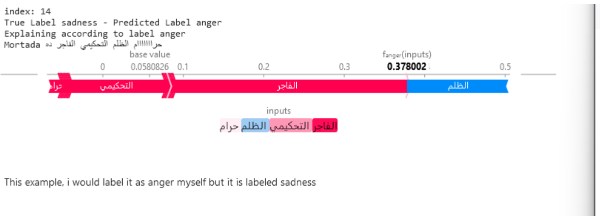
here it's labeled as fear and predicted as sadness however both labels here make sense and it's hard to pick one



```
index: 10
True Label fear - Predicted Label sadness
Explaining according to label sadness
يوجعني قلبي،لما اتمسك بانسان وانا خايفه يتركني،توجعني فكره ان الناس كلهم مؤقتين ومجرد عابرين.💔
73%  176/240 [00:00<00:00, 640.01it/s]
```

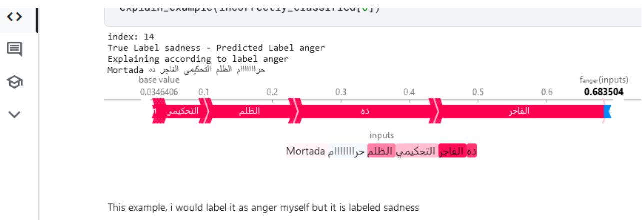
here it's labeled as fear and predicted as sadness however both labels here make sense and it's hard to pick one

Here the model made the correct prediction with light stemming.



```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada ده الفاجر التحكيمي الظلم حرااااااام
```

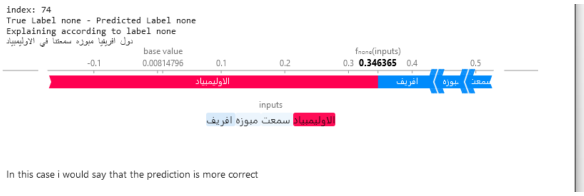
This example, i would label it as anger myself but it is labeled sadness



```
explain_example(incorrectly_classified[0])
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada ده الفاجر التحكيمي الظلم حرااااااام
```

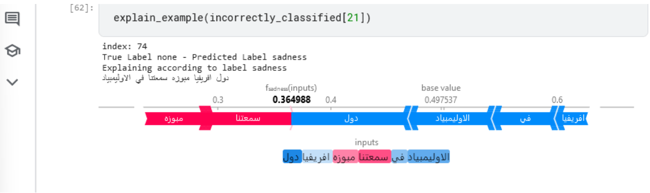
This example, i would label it as anger myself but it is labeled sadness

The model stopped using stopwords as explanations.



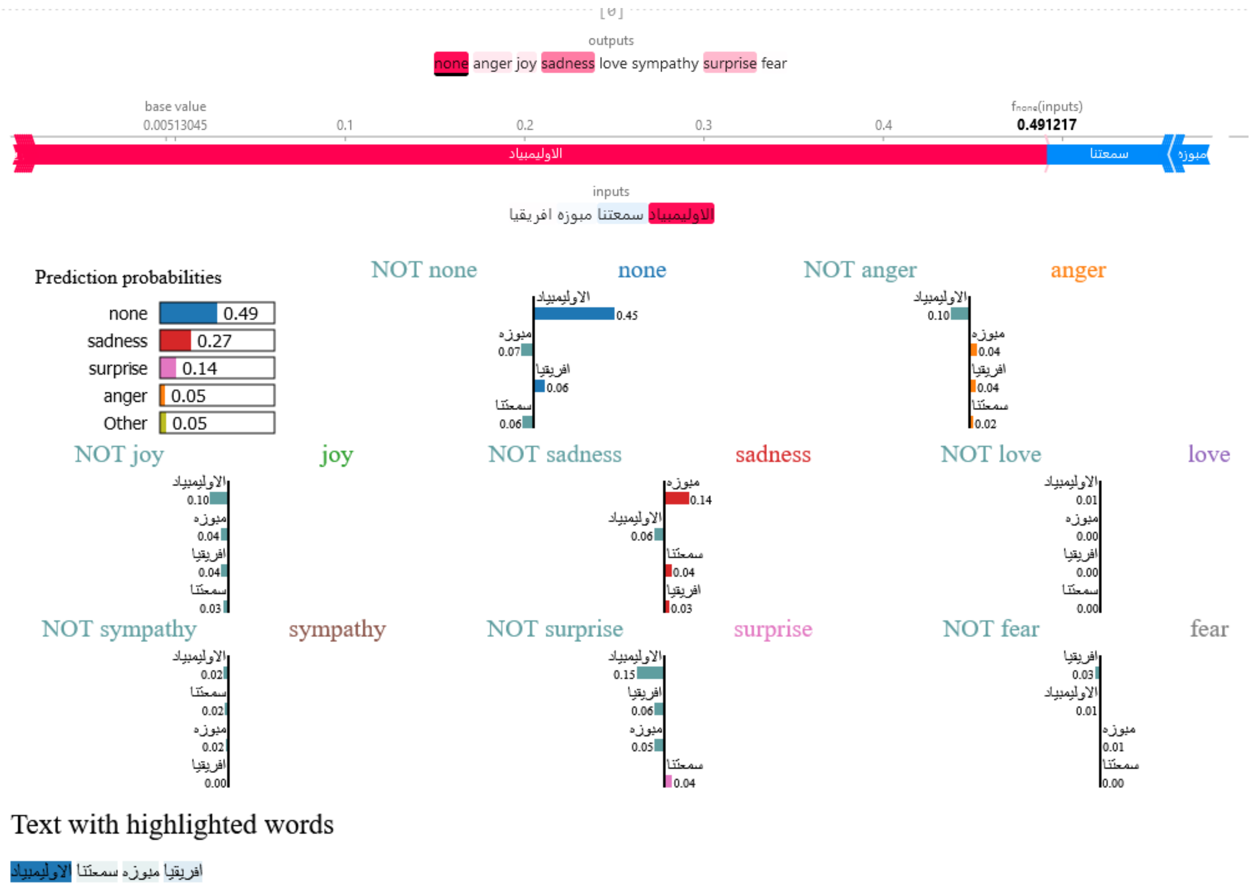
```
index: 74
True Label none - Predicted Label none
Explaining according to label none
دول افريقيا مبوزه سمعتنا في الاوليمبياد
```

In this case i would say that the prediction is more correct



```
[62]: explain_example(incorrectly_classified[21])
index: 74
True Label none - Predicted Label sadness
Explaining according to label sadness
دول افريقيا مبوزه سمعتنا في الاوليمبياد
```

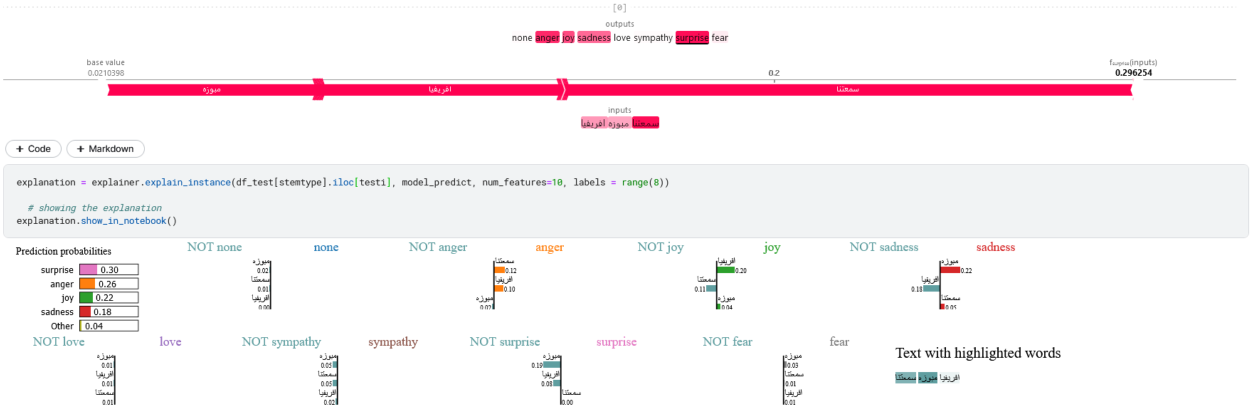
This was a case which I agreed with the prediction more than the label, but now the model got the label correctly. The explanation is bad too. ***(try the correct word)



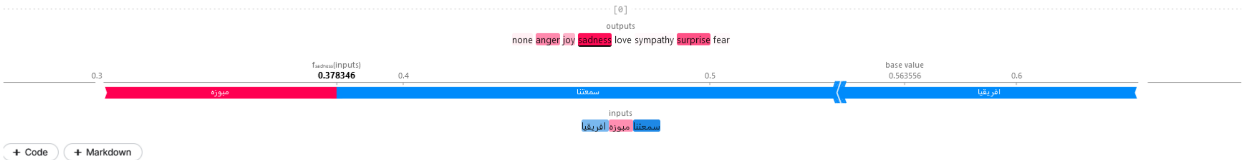
No significant change after correction?

Only that the word turned from being not-none to none but its score is not significant enough to change the outcome in either cases

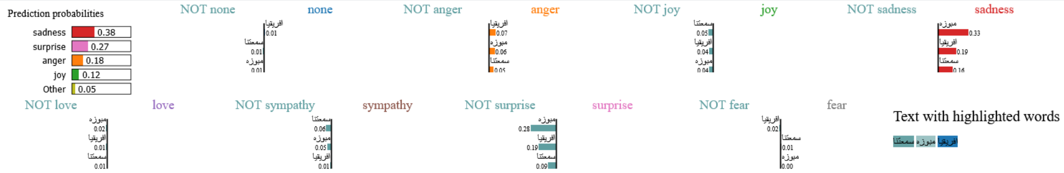
Let's us try without the problematic word:


```
explanation = explainer.explain_instance(df_test[stemtype].iloc[testi], model_predict, num_features=10, labels = range(8))

  # showing the explanation
explanation.show_in_notebook()
```

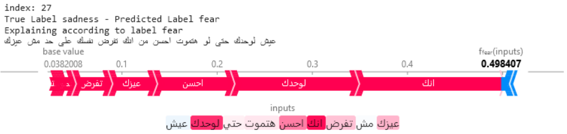
Now let us fix the word:



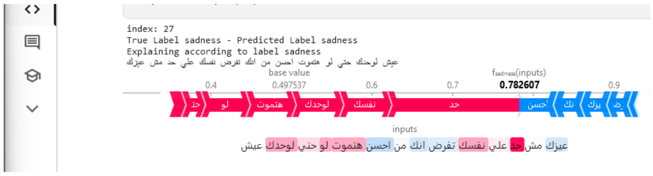
```
explanation = explainer.explain_instance(df_test[stemtype].iloc[testi], model_predict, num_features=10, labels = range(8))

  # showing the explanation
explanation.show_in_notebook()
```



We returned to sadness once more.

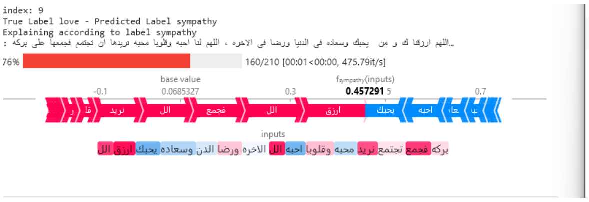


```
index: 27
True Label sadness - Predicted Label fear
Explaining according to label fear
عيش لوحدك حتى لو هتموت احسن من انك تفرض نفسك على حد مش عيزك
```

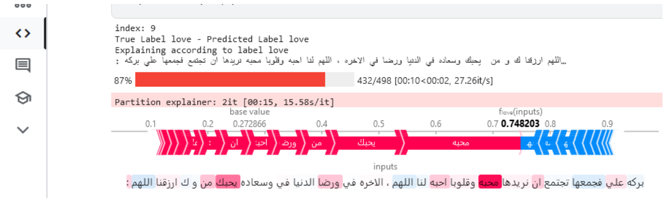


```
index: 27
True Label sadness - Predicted Label sadness
Explaining according to label sadness
عيش لوحدك حتي لو هتموت احسن من انك تفرض نفسك علي حد مش عيزك
```

This one just got more confusing.



```
index: 9
True Label love - Predicted Label sympathy
Explaining according to label sympathy
…اللهم ارزقنا ك و من  يحبك وسعاده فى الدنيا ورضا فى الاخره ، اللهم لنا احبه وقلوبا محبه تريدها ان تجتمع فجمعها على بركه :
76%                              160/210 [00:01<00:00, 475.79it/s]
```



```
index: 9
True Label love - Predicted Label love
Explaining according to label love
…اللهم ارزقنا ك و من  يحبك وسعاده في الدنيا ورضا في الاخره ، اللهم لنا احبه وقلوبا محبه تريدها ان تجتمع فجمعها علي بركه :
87%                              432/498 [00:10<00:02, 27.26it/s]
Partition explainer: 2it [00:15, 15.58s/it]
```

base value
0.1
0.2
0.272866
0.4
0.5
0.6
0.7
flove(inputs)
0.748203
0.8
0.9
ان
احب
ورض
من
يحبك
محبه
inputs
بركه علي فجمعها تجتمع ان تريدها محبه وقلوبا احبه لنا اللهم ، الاخره في ورضا الدنيا في وسعاده يحبك من و ك ارزقنا اللهم :

The model got confused in this case, as light stemming loses some of the context, so the mistake is understandable.

Conclusion:

The explanations got better but some of the context was lost.

Shap vs Lime

Interruption:

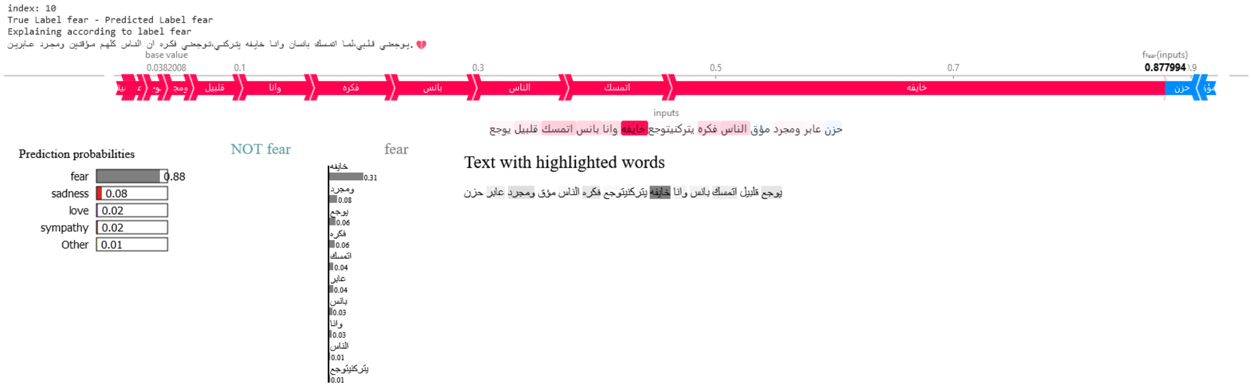
Some errors were found in light stemming during the comparison of lime and shap.

Raw data:

💔.يوجعني قلبي،لما اتمسك بانسان وانا خايفه يتركني،توجعني فكره ان الناس كلهم مؤقتين ومجرد عابرين

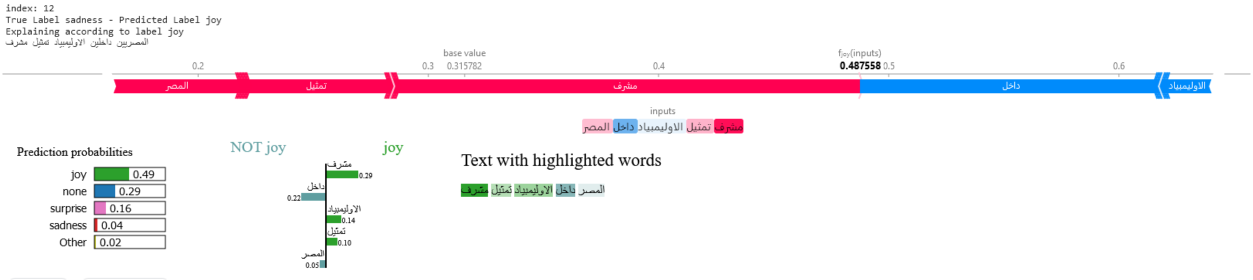
Light Stemming:

يوجع قلبيل اتمسك بانس وانا خايفه يتركنيتوجع فكره الناس مؤق ومجرد عابر حزن



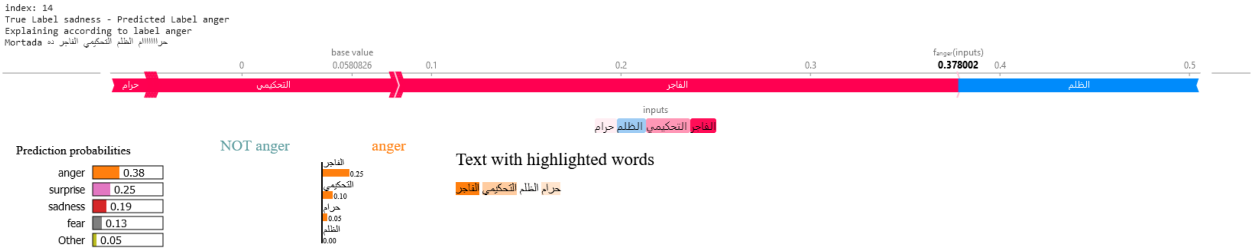
```
index: 10
True Label fear - Predicted Label fear
Explaining according to label fear
يوجعني قلبي،لما اتمسك باسان وانا خايفه يتركني،توجعني فكره ان الناس كلهم مؤقتين ومجرد عابرين.💔
```

While the main contributor to the label is the same, the other words are assigned different importance in different explainers.



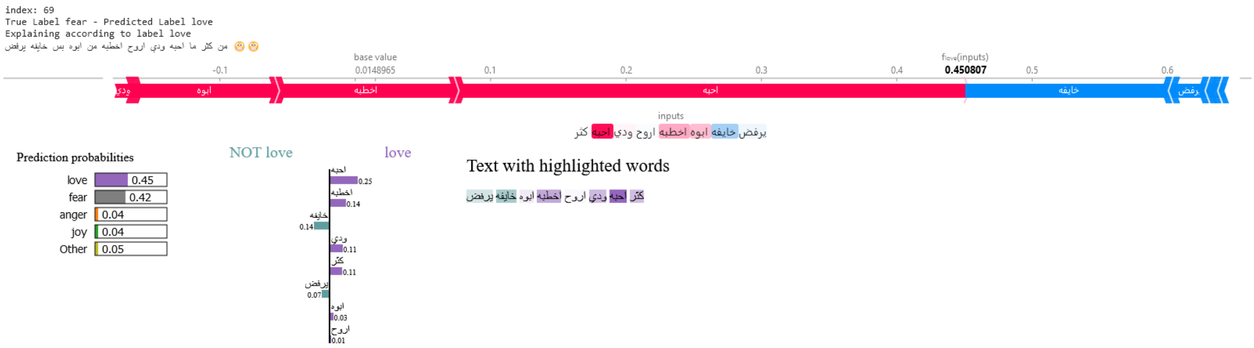
```
index: 12
True Label sadness - Predicted Label joy
Explaining according to label joy
المصريين داخلين الاوليمبياد تمثيل مشرف
```

The same words contribute to different labels according to different explainers.



```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada ده الفاجر التحكيمي الظلم حراااااام
```

Here, both have mostly the same explanation, but the words are assigned different weights.

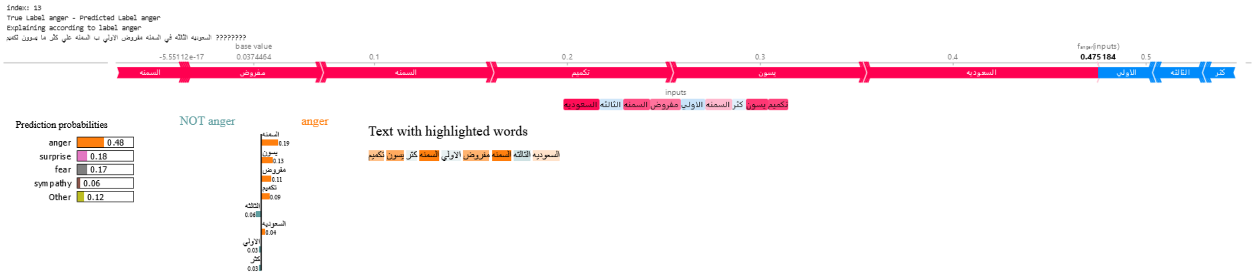


```
index: 69
True Label fear - Predicted Label love
Explaining according to label love
من كثر ما احبه ودي اروح اخطبه من ابوه بس خايفه يرفض 😬😬
```

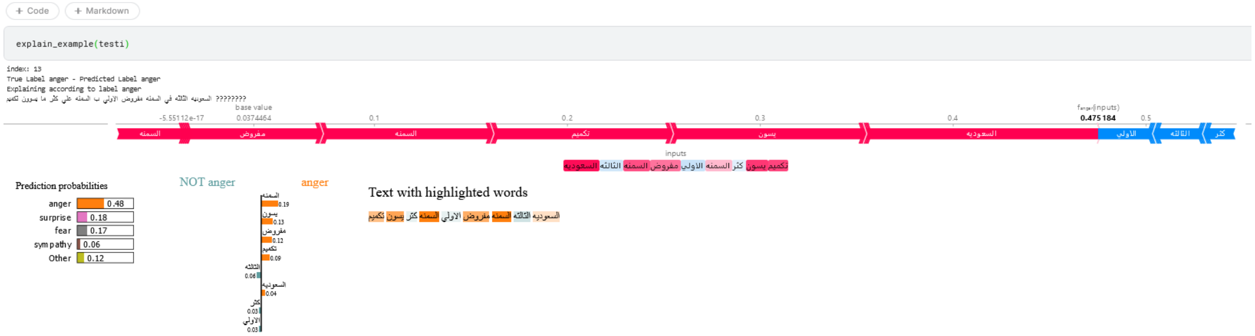
Here, LIME gave better insight into the workings of the model by providing the probabilities. We can see that even the model understands that there are mixed feelings.

Observe the weight of each words in the following examples:

1)



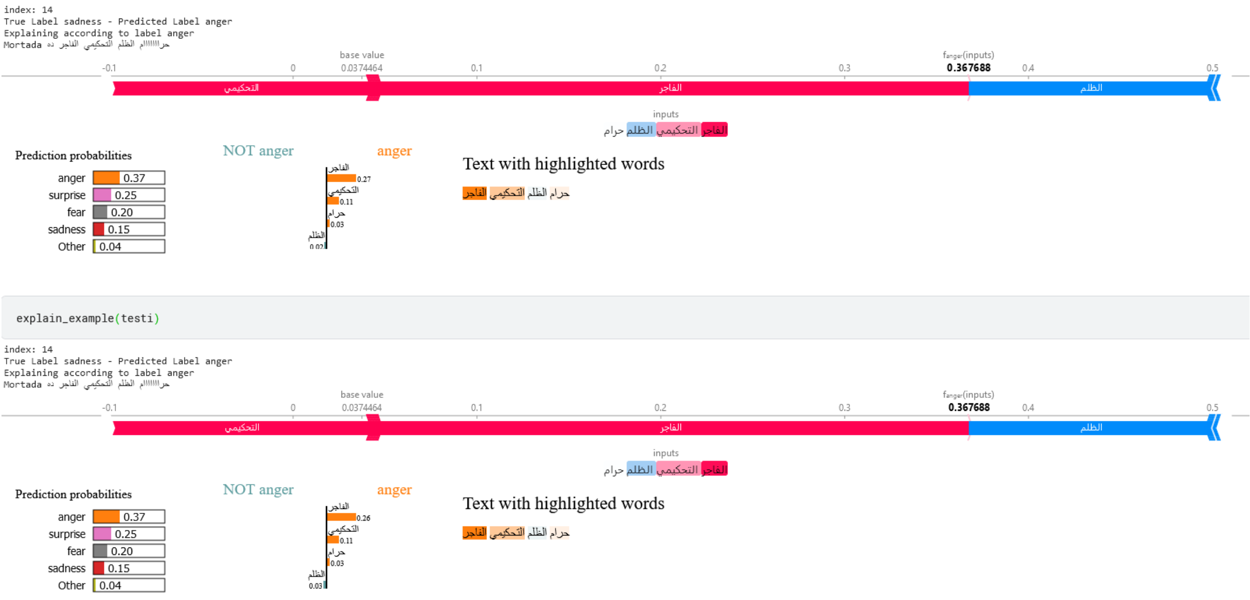
```
index: 13
True Label anger - Predicted Label anger
Explaining according to label anger
السعوديه الثالثه في السمنه مفروض الاولي ب السمنه علي كثر ما يسوون تكميم ????????
```



```
explain_example(testi)
```

```
index: 13
True Label anger - Predicted Label anger
Explaining according to label anger
السعوديه الثالثه في السمنه مفروض الاولي ب السمنه علي كثر ما يسوون تكميم ????????
```

2)



```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada حراااااام الظلم التحكيمي الفاجر ده
```

Text with highlighted words

حرام الظلم التحكيمي الفاجر

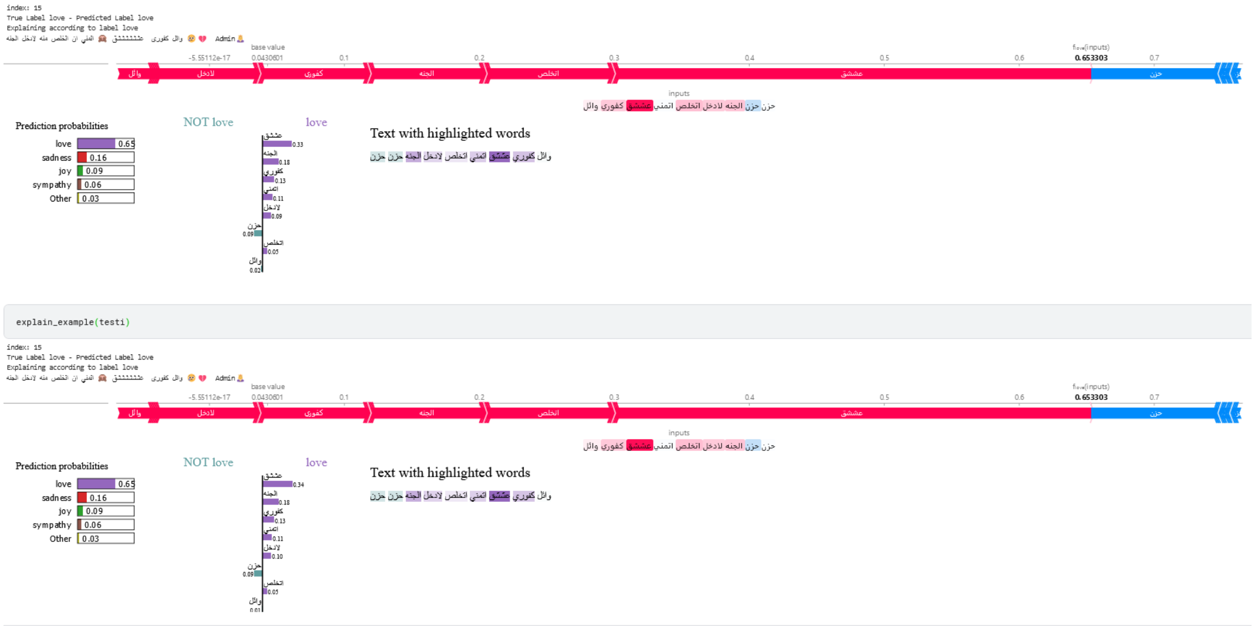
```
explain_example(testi)
```

```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada حراااااام الظلم التحكيمي الفاجر ده
```

Text with highlighted words

حرام الظلم التحكيمي الفاجر

3)



```
index: 15
True Label love - Predicted Label love
Explaining according to label love
Admin🙎 💔😢 وائل كفوري عشششششق 🙈 اتمني ان اتخلص منه لادخل الجنه
```

Text with highlighted words

وائل كفوري عشششق اتمني اتخلص لادخل الجنه حزن حزن

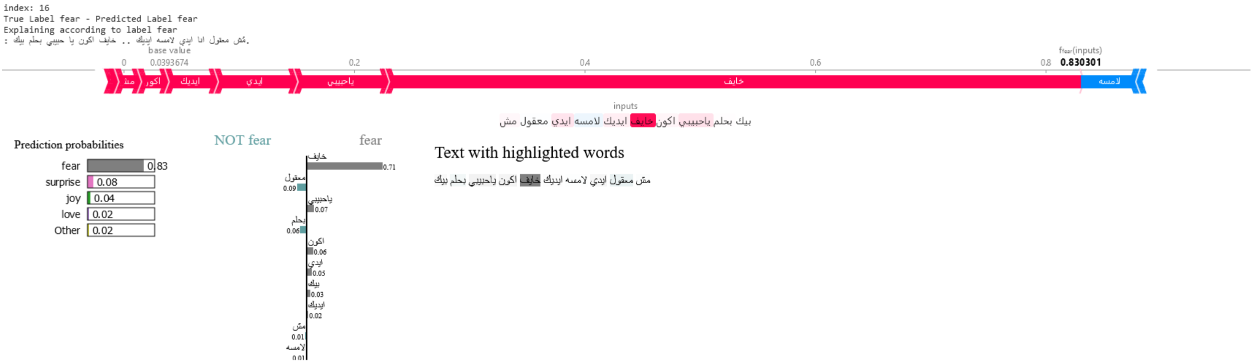
```
explain_example(testi)
```

```
index: 15
True Label love - Predicted Label love
Explaining according to label love
Admin🙎 💔😢 وائل كفوري عشششششق 🙈 اتمني ان اتخلص منه لادخل الجنه
```

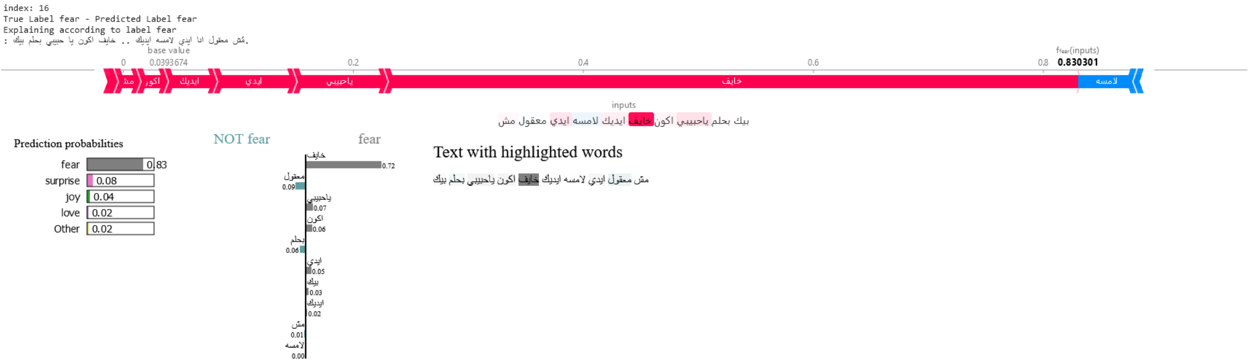
Text with highlighted words

وائل كفوري عشششق اتمني اتخلص لادخل الجنه حزن حزن

4)



```
index: 16
True Label fear - Predicted Label fear
Explaining according to label fear
: مُش معقول انا ايدي لامسه ايديك .. خايف اكون يا حبيبي بحلم بيك.
```

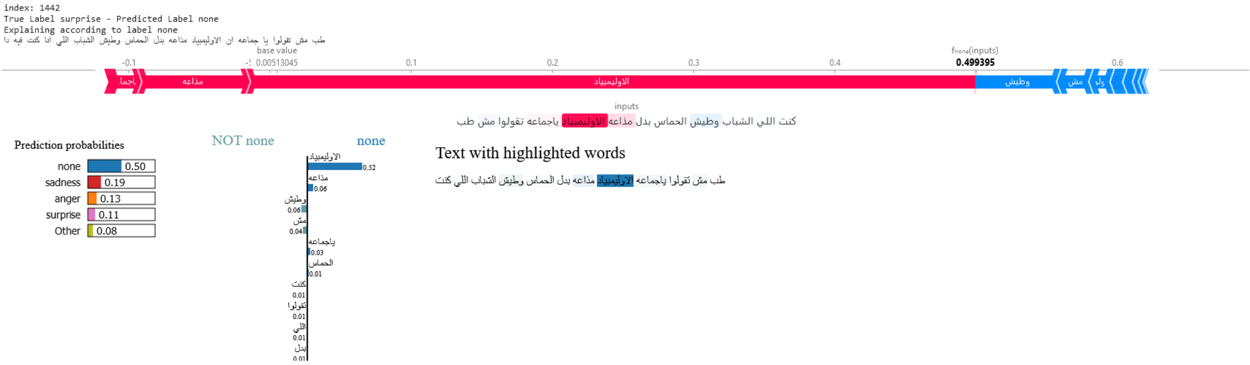


```
index: 16
True Label fear - Predicted Label fear
Explaining according to label fear
: مُش معقول انا ايدي لامسه ايديك .. خايف اكون يا حبيبي بحلم بيك.
```

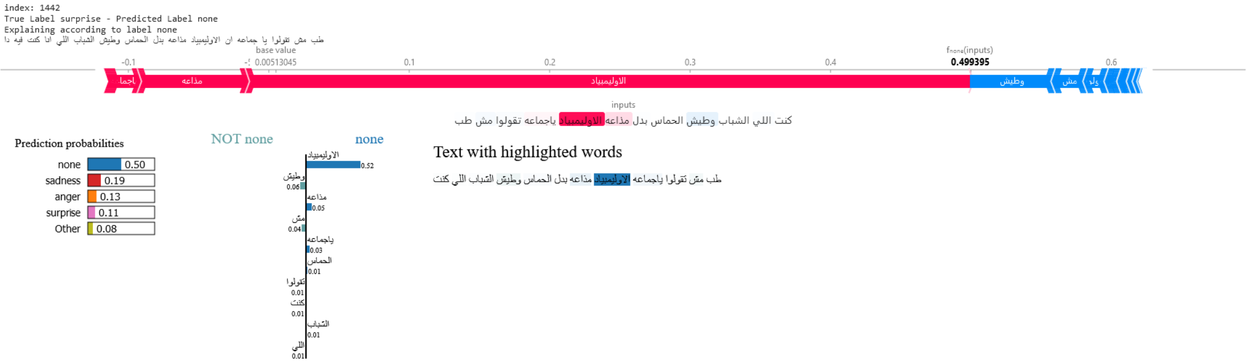
Due to the nature of its calculations, SHAP is very consistent in its output. Lime on the other hand is inconsistent due to its random nature.

Explaining the same example using Lime multiple times gives different weights each time but it’s usually generally correct.

The most significant word is almost always the same and the probabilities don’t appear to change but it would make a difference if the values were close to each other such that a 0.01 is enough to make a difference.



```
index: 1442
True Label surprise - Predicted Label none
Explaining according to label none
طب مش تقولوا يا جماعه ان الاوليمبياد مذاعه بدل الحماس وطيش الشباب اللي انا كنت فيه دا
```



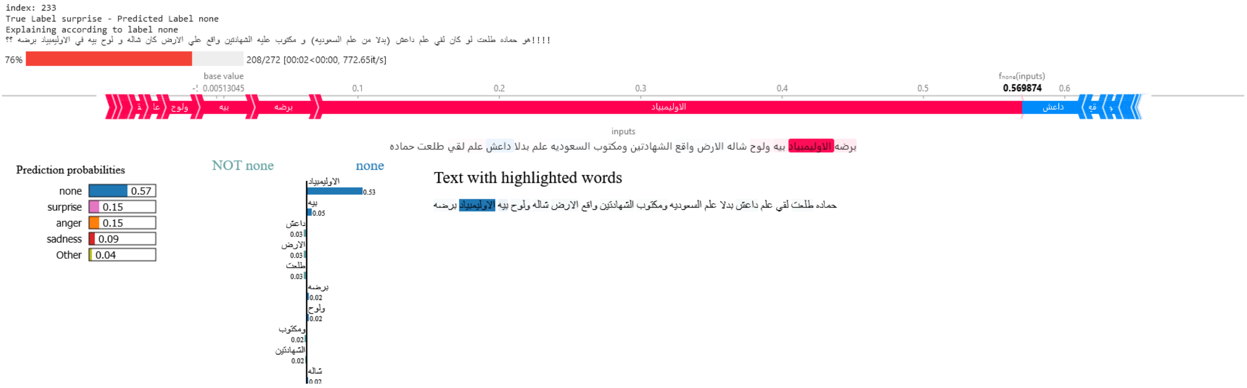
```
index: 1442
True Label surprise - Predicted Label none
Explaining according to label none
طب مش تقولوا يا جماعه ان الاولمبياد مذاعه بدل الحماس وطيش الشباب اللي انا كنت فيه دا
```

كنت اللي الشباب وطيش الحماس بدل مذاعه الاولمبياد ياجماعه تقولوا مش طب

Text with highlighted words

طب مش تقولوا ياجماعه الاولمبياد مذاعه بدل الحماس وطيش الشباب اللي كنت

Here the order of significance is different.

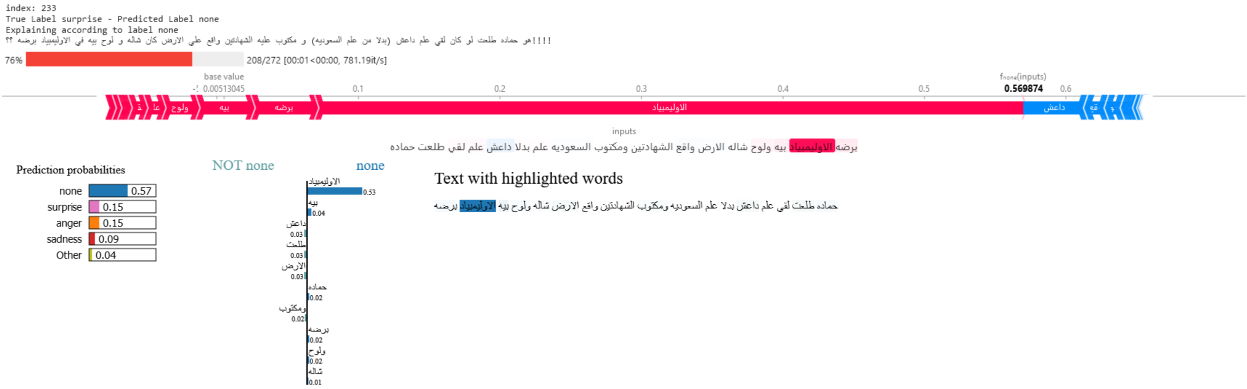


```
index: 233
True Label surprise - Predicted Label none
Explaining according to label none
!!!!هو حماده طلعت لو كان لقي علم داعش (بدلا من علم السعوديه) و مكتوب عليه الشهادتين واقع علي الارض كان شاله و لوح بيه في الاولمبياد برضه ؟؟
76% 208/272 [00:02<00:00, 772.65it/s]
```

برضه الاولمبياد بيه ولوح شاله الارض واقع الشهادتين ومكتوب السعوديه علم بدلا داعش علم لقي طلعت حماده

Text with highlighted words

حماده طلعت لقي علم داعش بدلا علم السعوديه ومكتوب الشهادتين واقع الارض شاله ولوح بيه الاولمبياد برضه



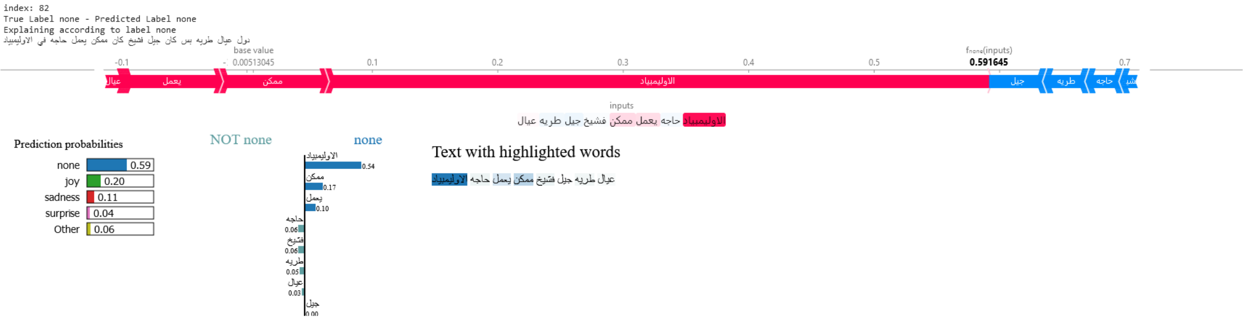
```
index: 233
True Label surprise - Predicted Label none
Explaining according to label none
!!!!هو حماده طلعت لو كان لقي علم داعش (بدلا من علم السعوديه) و مكتوب عليه الشهادتين واقع علي الارض كان شاله و لوح بيه في الاولمبياد برضه ؟؟
76% 208/272 [00:01<00:00, 781.19it/s]
```

برضه الاولمبياد بيه ولوح شاله الارض واقع الشهادتين ومكتوب السعوديه علم بدلا داعش علم لقي طلعت حماده

Text with highlighted words

حماده طلعت لقي علم داعش بدلا علم السعوديه ومكتوب الشهادتين واقع الارض شاله ولوح بيه الاولمبياد برضه

Here we can observe that some words appeared in the first sentence but not the second sentence and vice versa.

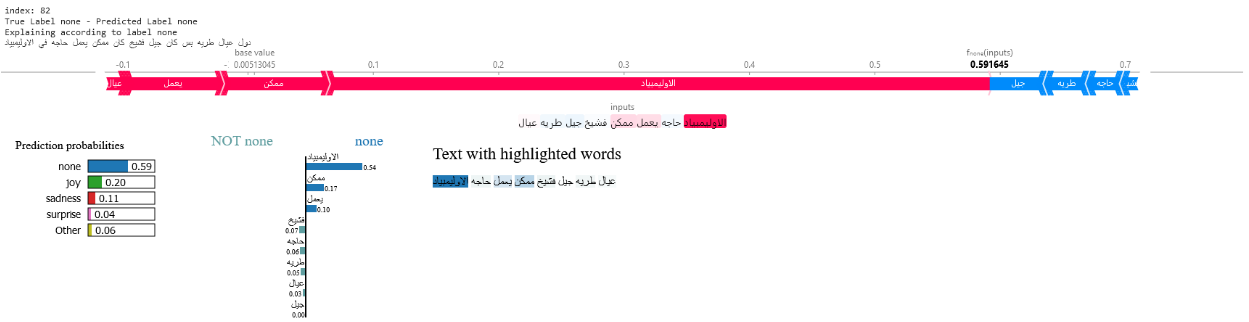


```
index: 82
True Label none - Predicted Label none
Explaining according to label none
دول عيال طريه بس كان جيل فشيخ كان ممكن يعمل حاجه في الاولمبياد
```

الاولمبياد حاجه يعمل ممكن فشيخ جيل طريه عيال

Text with highlighted words

عيال طريه جيل فشيخ ممكن يعمل حاجه الاولمبياد

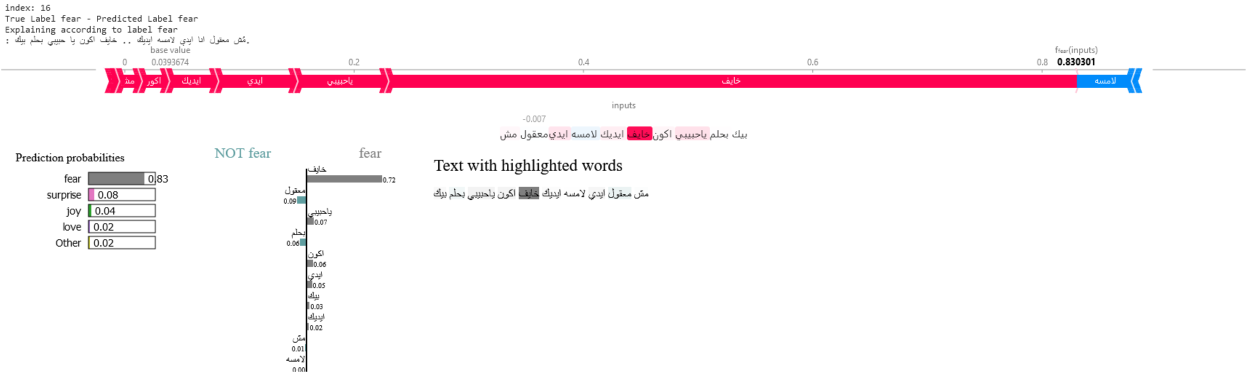


```
index: 82
True Label none - Predicted Label none
Explaining according to label none
دول عيال طريه بس كان جيل فشيخ كان ممكن يعمل حاجه في الاوليمبياد
```

As we can see here, a small change was enough to make a difference.

The last word switched from None to Not-none.

Now, we attempted to figure out why some words don't have weight in shap but have high weight in lime:

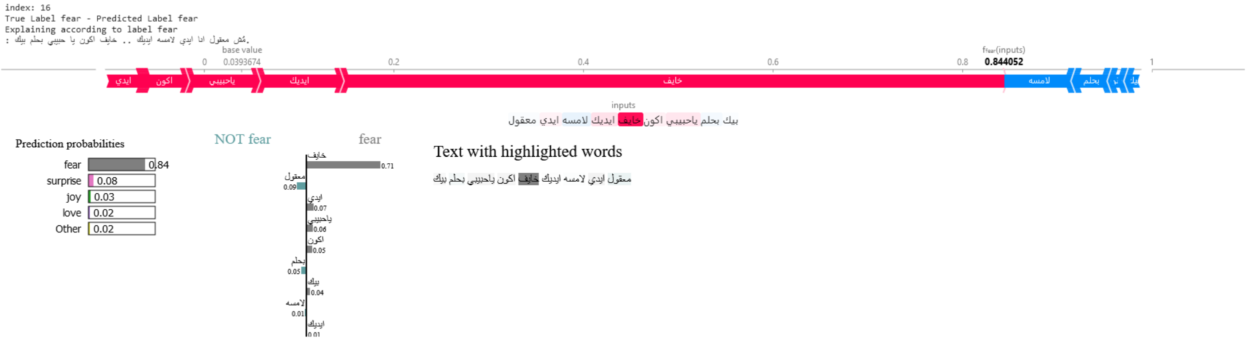


```
index: 16
True Label fear - Predicted Label fear
Explaining according to label fear
: مُش معقول انا ايدي لامسه ايديك .. خايف اكون يا حبيبي بحلم بيك.
```

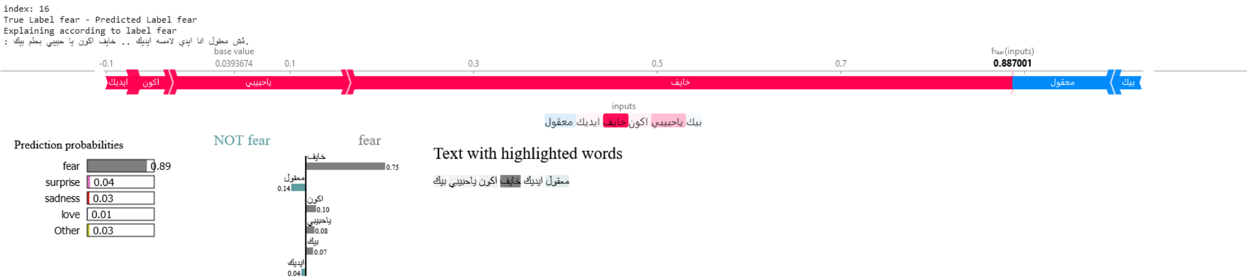
Let's return to this example once more.

The word 'معقول' has a weight of -0.09 in lime (it was verified by multiple runs) but appears to have no significant weight with shape (-0.007)

Across multiple runs I tried to remove words that has little to no weight in both lime and shap as following:



```
index: 16
True Label fear - Predicted Label fear
Explaining according to label fear
: مُش معقول انا ايدي لامسه ايديك .. خايف اكون يا حبيبي بحلم بيك.
```

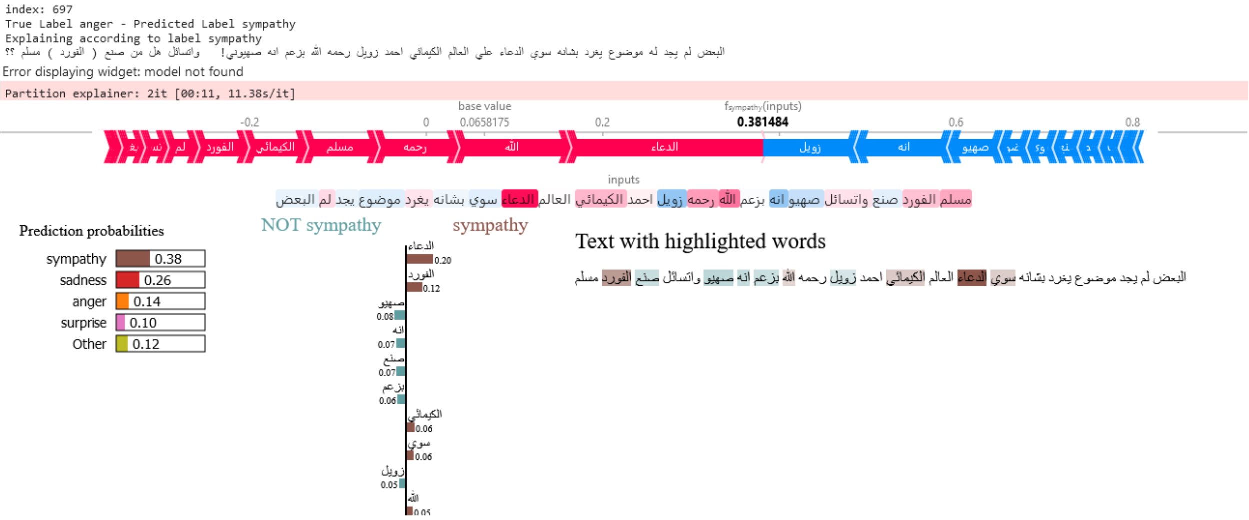


```
index: 16
True Label fear - Predicted Label fear
Explaining according to label fear
مُش معقول انا ايدي لامسه ايديك .. خايف اكون يا حبيبي بحلم بيك :
```

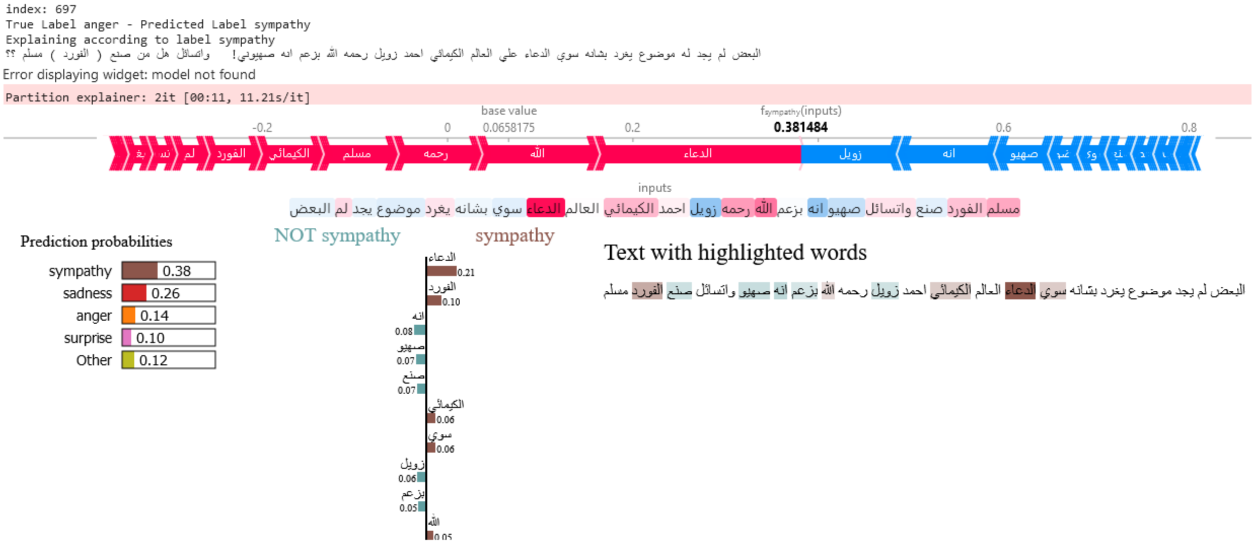
As we can observe in lime, other than the removed word, the words have the same order and a little difference in weight.

But in Shap, each removed word has a significant effect on the rest of the weights even if the removed word's weight is not significant itself.

Observe this 21-words example.



```
index: 697
True Label anger - Predicted Label sympathy
Explaining according to label sympathy
البعض لم يجد له موضوع يغرد بشانه سوي الدعاء علي العالم الكيمائي احمد زويل رحمه الله بزعم انه صهيوني!   واتسائل هل من صنع ( الفورد ) مسلم ؟؟
Error displaying widget: model not found
Partition explainer: 2it [00:11, 11.38s/it]
```



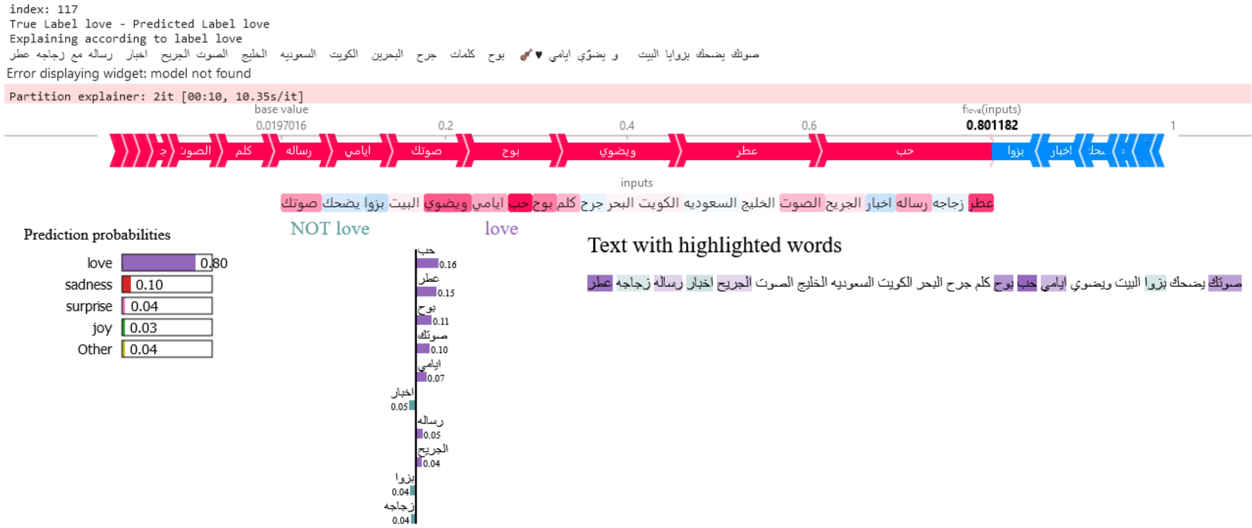
```
index: 697
True Label anger - Predicted Label sympathy
Explaining according to label sympathy
البعض لم يجد له موضوع يغرد بشانه سوي الدعاء علي العالم الكيمائي احمد زويل رحمه الله بزعم انه صهيوني!   واتسائل هل من صنع ( الفورد ) مسلم ؟؟
```

Error displaying widget: model not found

```
Partition explainer: 2it [00:11, 11.21s/it]
```

البعض لم يجد له موضوع يغرد بشانه سوي الدعاء العالم الكيمائي احمد زويل رحمه الله بزعم انه صهيوني واتسائل صنع الفورد مسلم

2)



```
index: 117
True Label love - Predicted Label love
Explaining according to label love
صوتك يضحك بزوايا البيت  و يضوّي ايامي ♥🎻  بوح  كلمات  جرح  البحرين  الكويت  السعوديه  الخليج  الصوت الجريح  اخبار  رساله مع زجاجه عطر
```

Error displaying widget: model not found

```
Partition explainer: 2it [00:10, 10.35s/it]
```

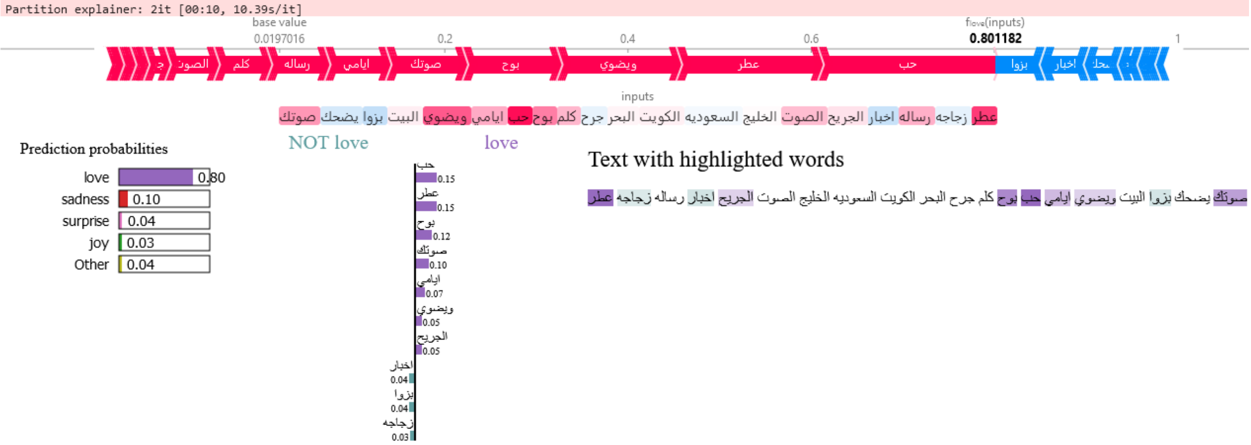
صوتك يضحك بزوا البيت ويضوي ايامي حب بوح كلم جرح البحر الكويت السعوديه الخليج الصوت الجريح اخبار رساله زجاجه عطر

```
index: 117
True Label love - Predicted Label love
Explaining according to label love
```

صوتك يضحك بزوايا البيت و يضوّي ايامي ♥🎻 بوح كلمات جرح البحرين الكويت السعوديه الخليج الصوت الجريح اخبار رساله مع زجاجه عطر

Error displaying widget: model not found

```
Partition explainer: 2it [00:10, 10.39s/it]
```



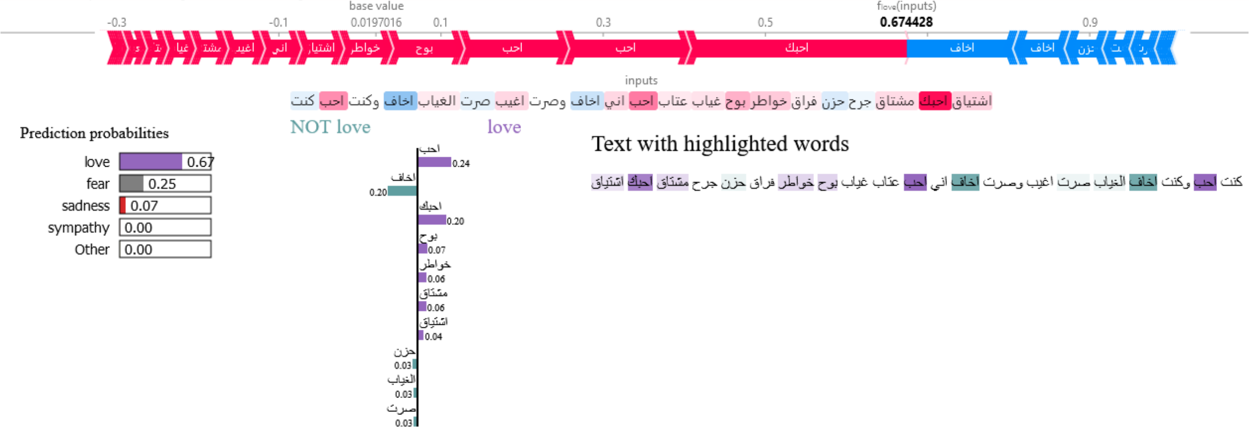
3)

```
index: 838
True Label love - Predicted Label love
Explaining according to label love
```

كنت احب وكنت اخاف من الغياب صرت اغيب وصرت اخاف اني احب عتاب غياب بوح خواطر فراق حزن وله جرح مشتاق احبك اشتياق

Error displaying widget: model not found

```
Partition explainer: 2it [00:11, 11.62s/it]
```

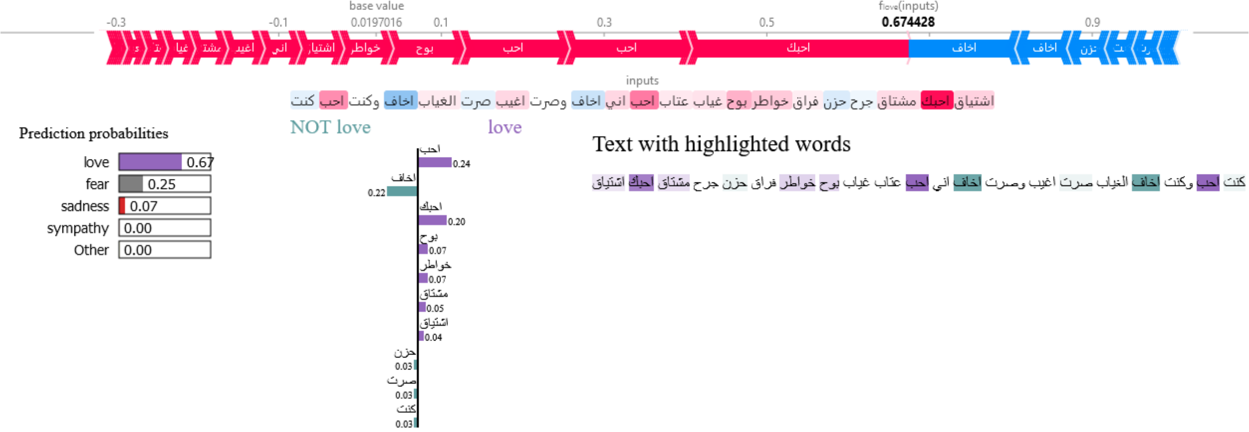


```
index: 838
True Label love - Predicted Label love
Explaining according to label love
```

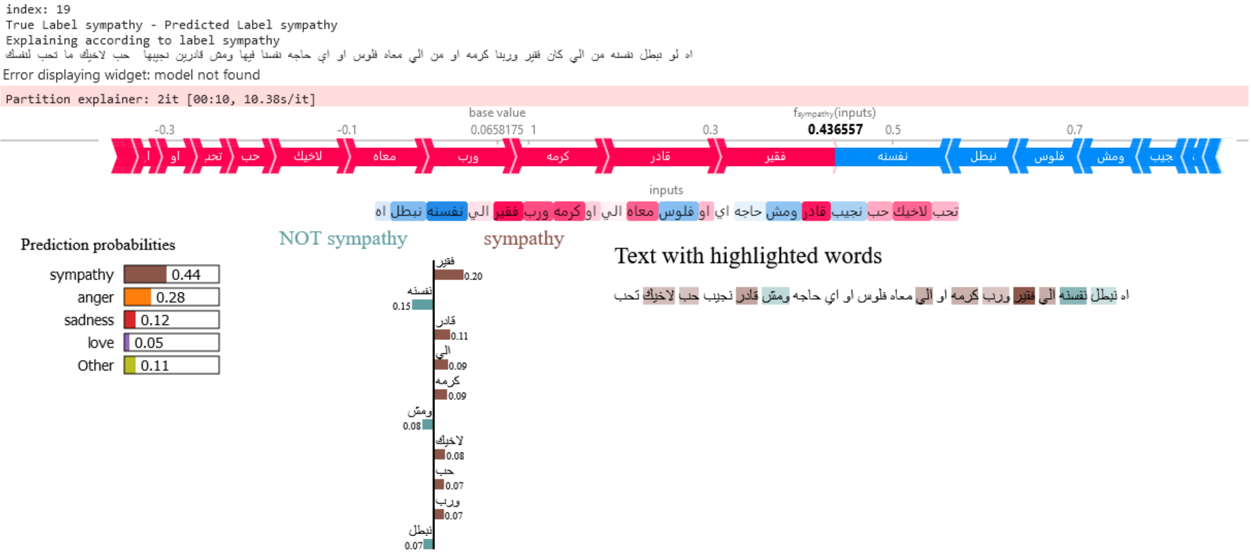
كنت احب وكنت اخاف من الغياب صرت اغيب وصرت اخاف اني احب عتاب غياب بوح خواطر فراق حزن وله جرح مشتاق احبك اشتياق

Error displaying widget: model not found

```
Partition explainer: 2it [00:11, 11.36s/it]
```



4)



```
index: 19
True Label sympathy - Predicted Label sympathy
Explaining according to label sympathy
اه لو نبطل نفسنه من الي كان فقير وربنا كرمه او من الي معاه فلوس او اي حاجه نفسنا فيها ومش قادرين نجيبها  حب لاخيك ما تحب لنفسك
```

Error displaying widget: model not found

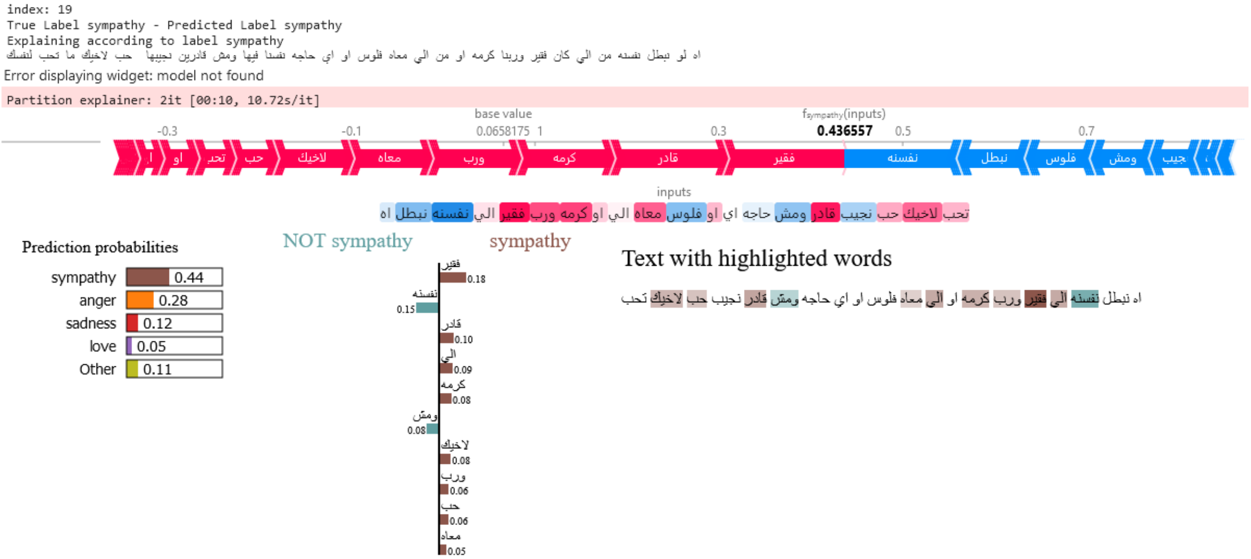
```
Partition explainer: 2it [00:10, 10.38s/it]
```

Prediction probabilities

sympathy	0.44
anger	0.28
sadness	0.12
love	0.05
Other	0.11

Text with highlighted words

اه نبطل نفسنه الي فقير ورب كرمه او الي معاه فلوس او اي حاجه ومش قادر نجيب حب لاخيك تحب



```
index: 19
True Label sympathy - Predicted Label sympathy
Explaining according to label sympathy
اه لو نبطل نفسنه من الي كان فقير وربنا كرمه او من الي معاه فلوس او اي حاجه نفسنا فيها ومش قادرين نجيبها  حب لاخيك ما تحب لنفسك
```

Error displaying widget: model not found

```
Partition explainer: 2it [00:10, 10.72s/it]
```

Prediction probabilities

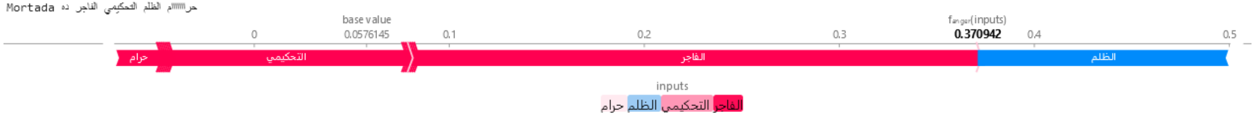
sympathy	0.44
anger	0.28
sadness	0.12
love	0.05
Other	0.11

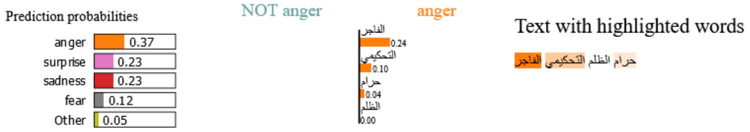
Text with highlighted words

اه نبطل نفسنه الي فقير ورب كرمه او الي معاه فلوس او اي حاجه ومش قادر نجيب حب لاخيك تحب

Let’s observe small examples now:

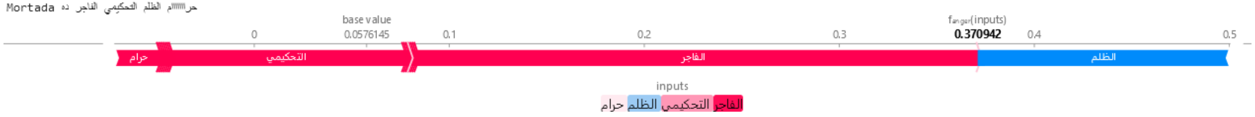
```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada حراااااام الظلم التحكيمي الفاجر ده
```

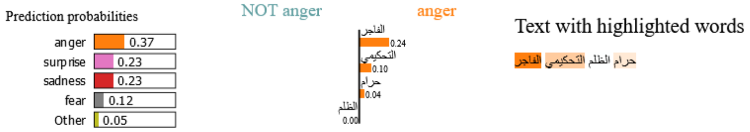




```
explain_example(testi)
```

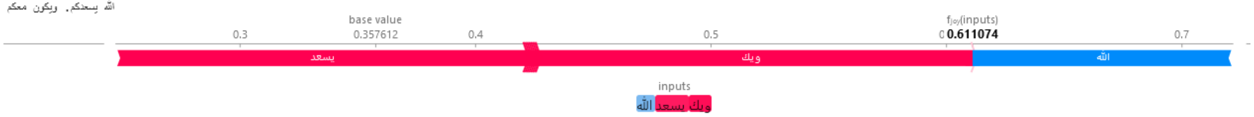
```
index: 14
True Label sadness - Predicted Label anger
Explaining according to label anger
Mortada حراااااام الظلم التحكيمي الفاجر ده
```

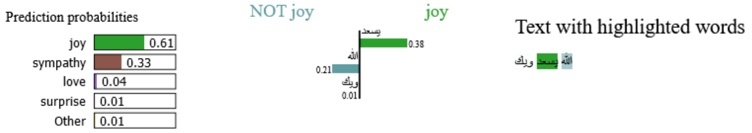




2)

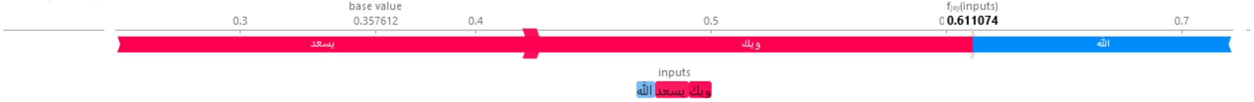
```
index: 36
True Label sympathy - Predicted Label joy
Explaining according to label joy
الله يسعدكم. ويكون معكم
```

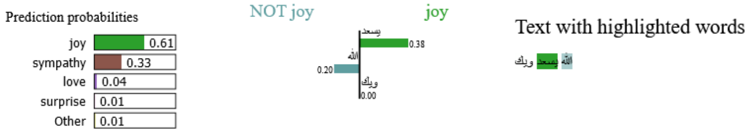




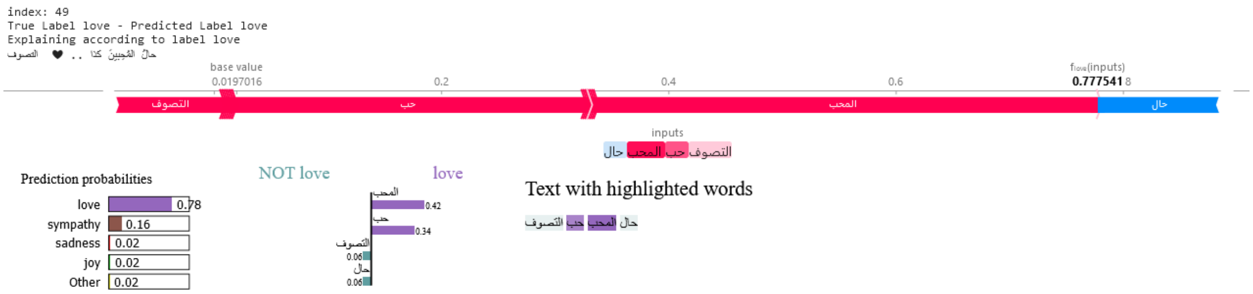
```
explain_example(testi)
```

```
index: 36
True Label sympathy - Predicted Label joy
Explaining according to label joy
الله يسعدكم. ويكون معكم
```

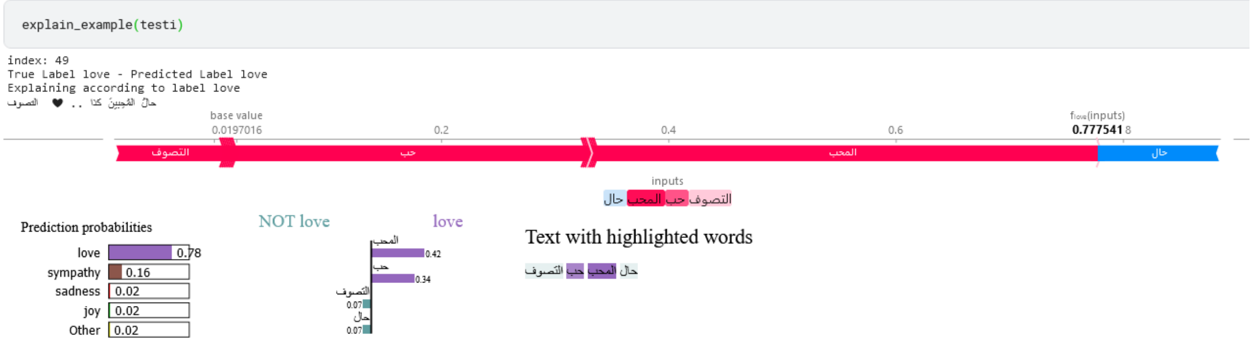




3)



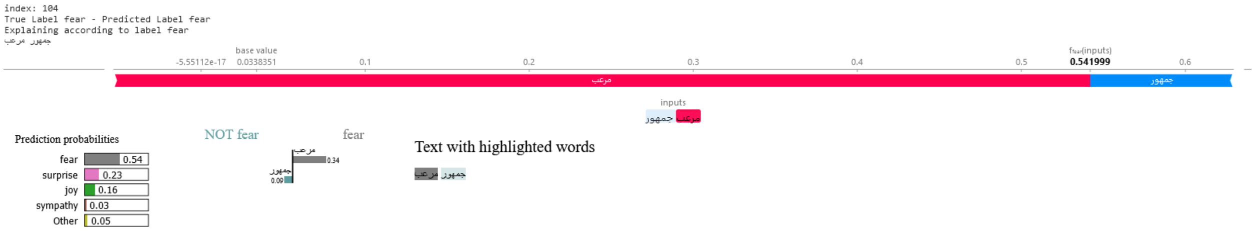
```
index: 49
True Label love - Predicted Label love
Explaining according to label love
حالُ المُحِبينَ كذا .. ❤ التصوف
```



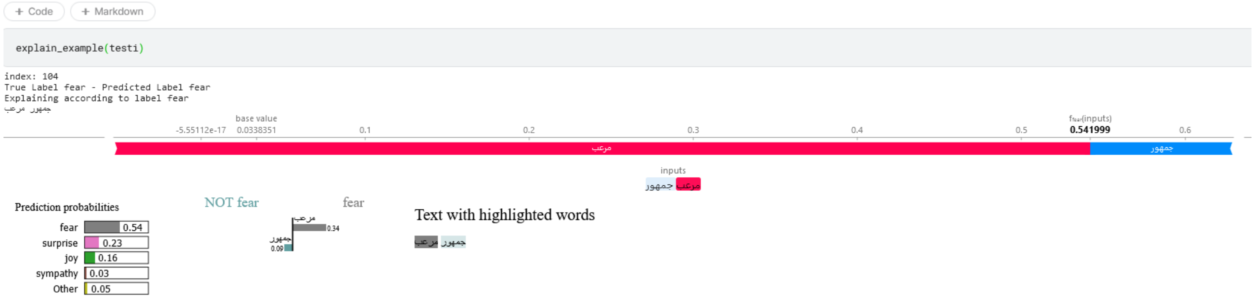
```
explain_example(testi)
```

```
index: 49
True Label love - Predicted Label love
Explaining according to label love
حالُ المُحِبينَ كذا .. ❤ التصوف
```

4)



```
index: 104
True Label fear - Predicted Label fear
Explaining according to label fear
جمهور مرعب
```



```
explain_example(testi)
```

```
index: 104
True Label fear - Predicted Label fear
Explaining according to label fear
جمهور مرعب
```

It seems that lime is unable to explain it properly and It may be because Lime generates 5000 random samples and while those examples are enough to cover most sentences, it appears that it is not enough to cover all permutations possible in a long sentence so, the changes between different runs are a bit higher.

The highest change in short examples tend to be 0.01 but in long sentences it can reach 0.02 of change in multiple words at once.

Conclusion:

Due to the way of calculation, Lime has little in the way of stability when running the same example multiple times but is surprisingly robust when some of the input is changed or dropped.

Shap on the other hand is very consistent in its calculations when running the same example, but it is greatly affected by small changes in the input and seemingly unrelated words may have significant changes in the output.

Interruption (again):

Some errors were found in light stemming during the comparison of lime and shap.

Raw data:

دول افريفيا مبوزه سمعتنا في الاولمبياد

Light Stemming:

افريف مبوزه سمعت الاولمبياد

There is a spelling mistake here, and some removed stopwords that need to be looked into.

And another one:

Raw data:

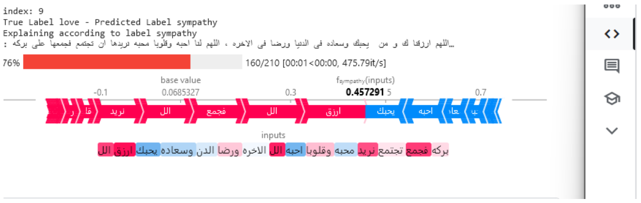
محدش عارف هما تعبوا اد ايه عشان يوصلوا الاوليمبياد و هم مش ناقصين تعليقات الجهله في اللعبه لادائهم و لا تعليقات المصريين
عموما علي لبسهم

Light Stemming:

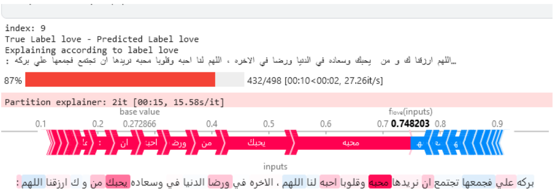
محدش عارف تعب اد عشان يوصل الاوليمبياد مش ناقص تعليق الجهله اللعبه لاداىٔ تعليق المصر عمو لبس

There are some removed stopwords that need to be looked into.

Using Cleaned data without stemming:



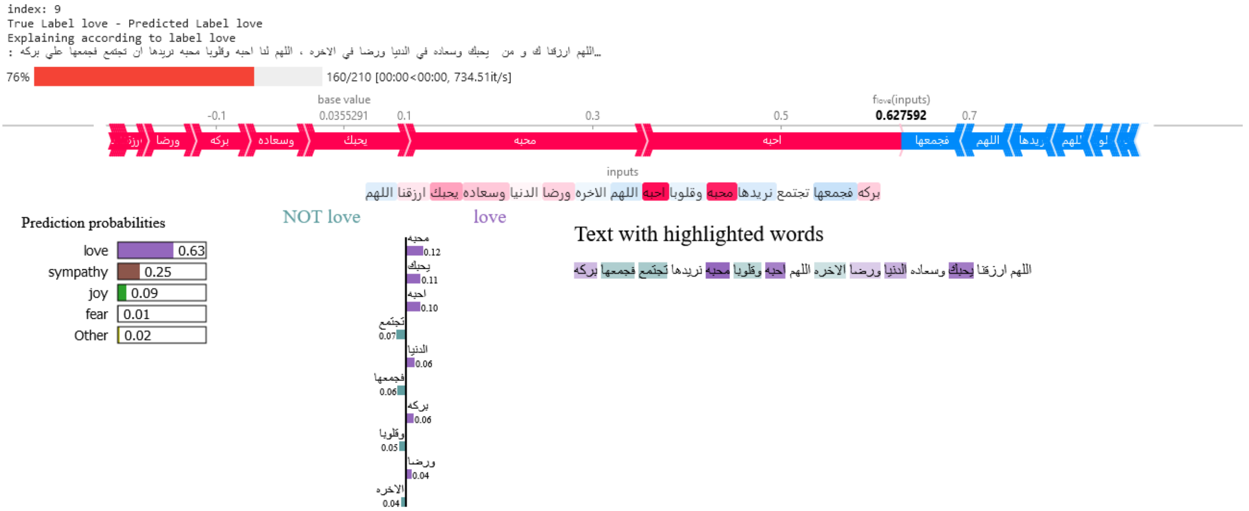
```
index: 9
True Label love - Predicted Label sympathy
Explaining according to label sympathy
اللهم ارزقنا ك و من  يحبك وسعاده فى الدنيا ورضا فى الاخره ، اللهم لنا احبه وقلوبا محبه تريدها ان تجتمع فجمعها على بركه :...
76% 160/210 [00:01<00:00, 475.79it/s]
```



```
index: 9
True Label love - Predicted Label love
Explaining according to label love
اللهم ارزقنا ك و من  يحبك وسعاده في الدنيا ورضا في الاخره ، اللهم لنا احبه وقلوبا محبه تريدها ان تجتمع فجمعها علي بركه :...
87% 432/498 [00:10<00:02, 27.26it/s]
Partition explainer: 2it [00:15, 15.58s/it]
```

base value
0.272866
f_{love}(inputs)
0.748203
0.1
0.2
0.4
0.5
0.6
0.7
0.8
0.9
ان
احبه
ورض
من
يحبك
محبه
inputs
بركه علي فجمعها تجتمع ان نريدها محبه وقلوبا احبه لنا اللهم ، الاخره في ورضا الدنيا في وسعاده يحبك من و ك ارزقنا اللهم :

There was this error that I attributed to a loss of context but without stemming, the model got it correctly once more:



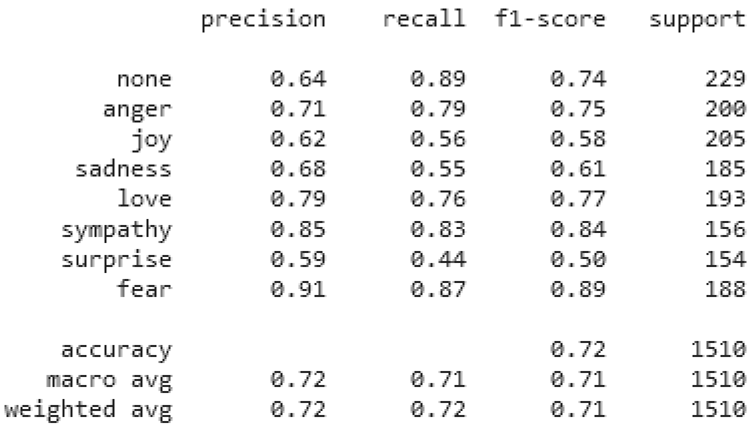
```
index: 9
True Label love - Predicted Label love
Explaining according to label love
اللهم ارزقنا ك و من  يحبك وسعاده في الدنيا ورضا في الاخره ، اللهم لنا احبه وقلوبا محبه تريدها ان تجتمع فجمعها علي بركه :...
76% 160/210 [00:00<00:00, 734.51it/s]
```

Text with highlighted words

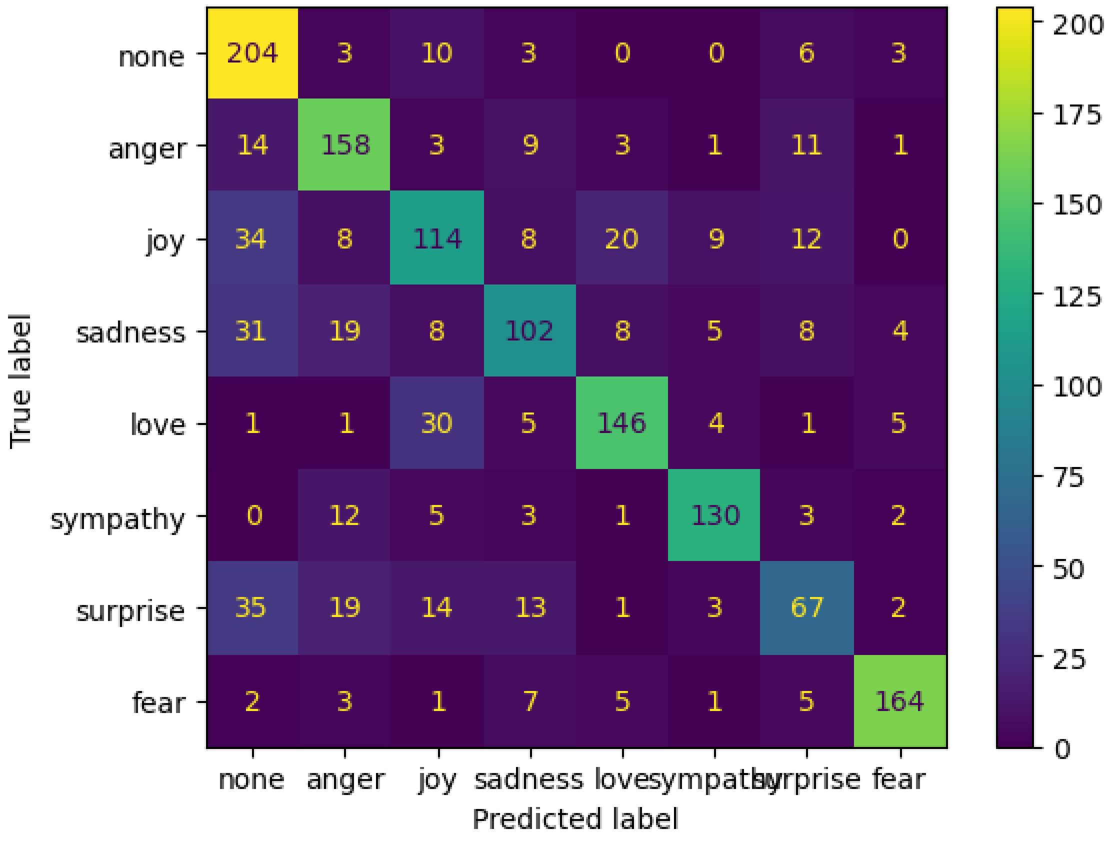
اللهم ارزقنا يحبك وسعاده الدنيا ورضا الاخره اللهم احبه وقلوبا محبه تريدها تجتمع فجمعها بركه

So, context can be lost with stemming but it appears that without stemming, it may get better.

Since we will be using cleaned data now, here are the metrics:



```
              precision    recall  f1-score   support

        none       0.64      0.89      0.74       229
       anger       0.71      0.79      0.75       200
         joy       0.62      0.56      0.58       205
     sadness       0.68      0.55      0.61       185
        love       0.79      0.76      0.77       193
    sympathy       0.85      0.83      0.84       156
    surprise       0.59      0.44      0.50       154
        fear       0.91      0.87      0.89       188

    accuracy                           0.72      1510
   macro avg       0.72      0.71      0.71      1510
weighted avg       0.72      0.72      0.71      1510
```

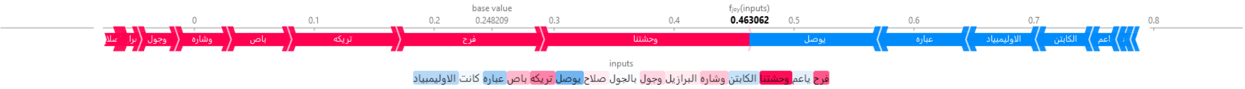


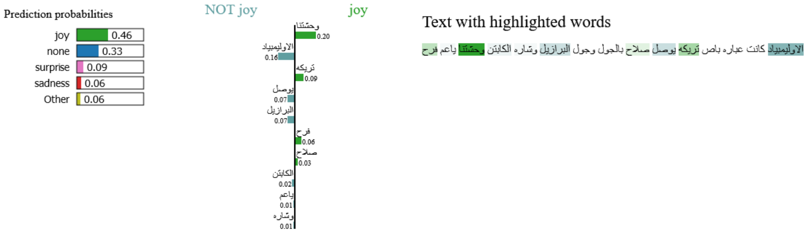
	none	anger	joy	sadness	love	sympathy	surprise	fear
none	204	3	10	3	0	0	6	3
anger	14	158	3	9	3	1	11	1
joy	34	8	114	8	20	9	12	0
sadness	31	19	8	102	8	5	8	4
love	1	1	30	5	146	4	1	5
sympathy	0	12	5	3	1	130	3	2
surprise	35	19	14	13	1	3	67	2
fear	2	3	1	7	5	1	5	164

By comparing the confusion matrix with the previous one, we can observe that the box that indicates (true = none, pred = joy) increased from 6 to 10.

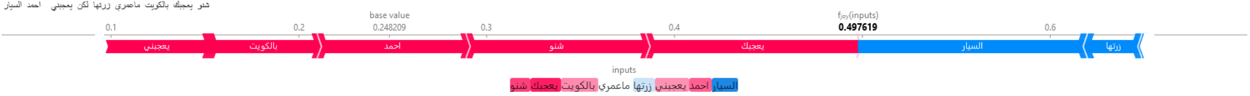
That was despite the fact that context appears to be more complete without stemming so lets examine those 10 samples.

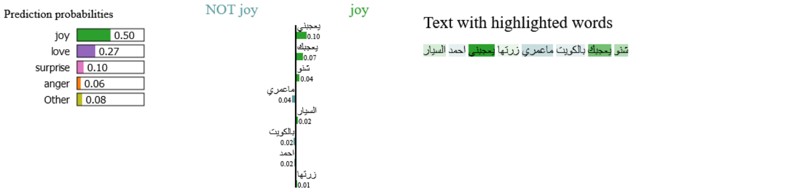
```
index: 26
True Label none - Predicted Label joy
Explaining according to label joy
الاوليمبياد كانت عباره عن باص من تريكه يوصل صلاح بالجول وجول البرازيل وشاره الكابتن .. وحشتنا ياعم :))
76% 160/210 [00:00<00:00, 738.97it/s]
```



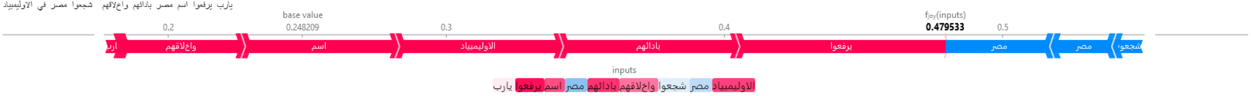


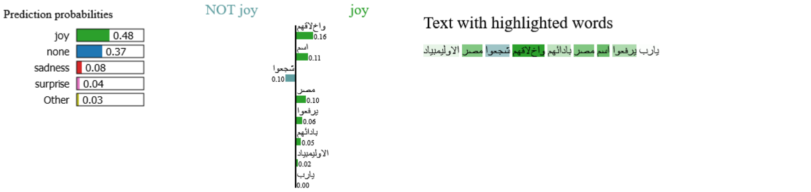
```
index: 100
True Label none - Predicted Label joy
Explaining according to label joy
شنو يعجبك بالكويت ماعمري زرتها لكن يعجبني  احمد السيار
```



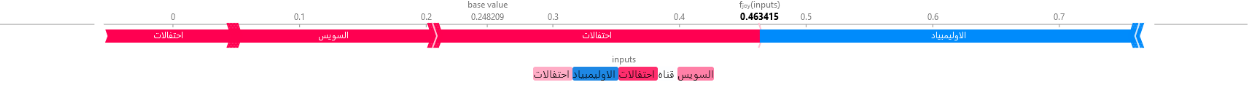


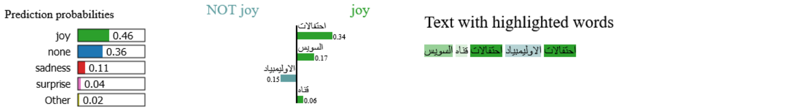
```
index: 325
True Label none - Predicted Label joy
Explaining according to label joy
يارب يرفعوا اسم مصر بادائهم واخلاقهم  شجعوا مصر في الاوليمبياد
```



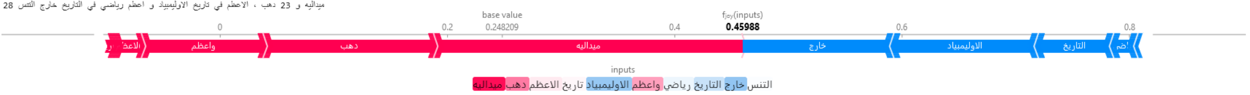


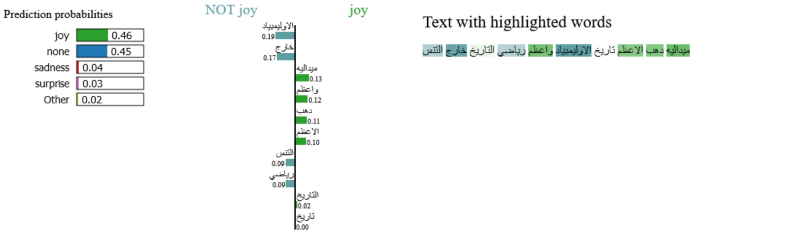
```
index: 801
True Label none - Predicted Label joy
Explaining according to label joy
0 احتفالات الاوليمبياد 1 احتفالات قناه السويس
```



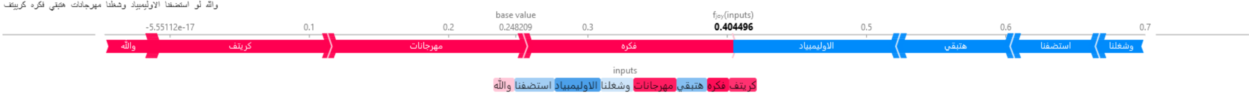


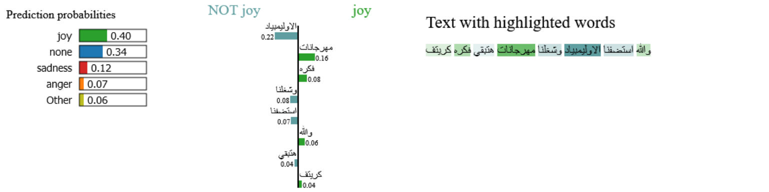
```
index: 815
True Label none - Predicted Label joy
Explaining according to label joy
ميداليه و 23 ذهب ، الاعظم في تاريخ الاوليمبياد و اعظم رياضي في التاريخ خارج التنس 28
```



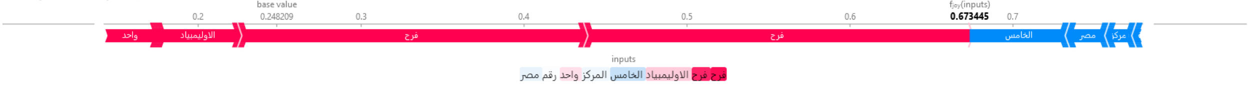


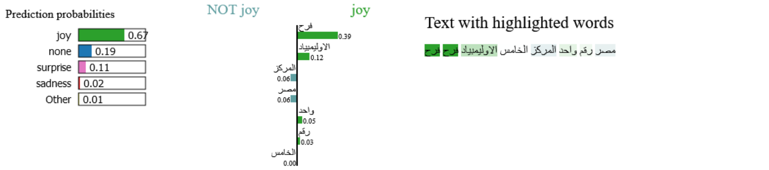
```
index: 874
True Label none - Predicted Label joy
Explaining according to label joy
والله لو استضفنا الاوليمبياد وشغلنا مهرجانات هتبقي فكره كريتف
```



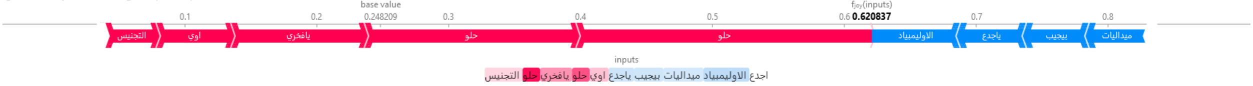


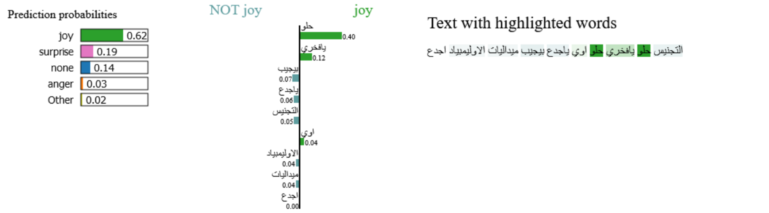
```
index: 904
True Label none - Predicted Label joy
Explaining according to label joy
مصر رقم واحد في المركز الخامس في الاوليمبياد 😀😀
```



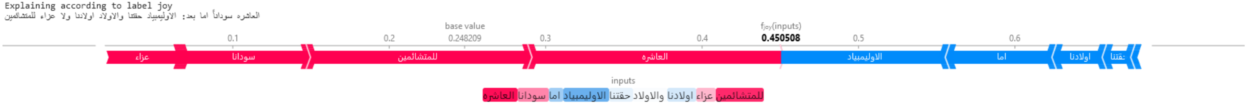


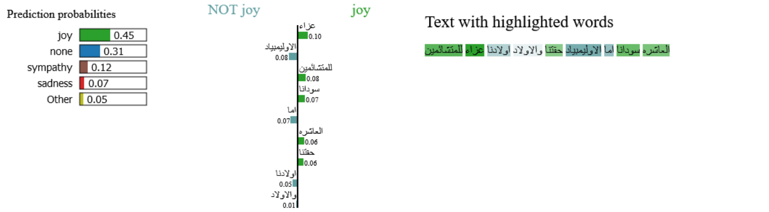
```
index: 917
True Label none - Predicted Label joy
Explaining according to label joy
Rio2016 التجنيس ده حلو يا فخري حلو اوي يا جدع بيجيب ميداليات في الاوليمبياد اجدع
```



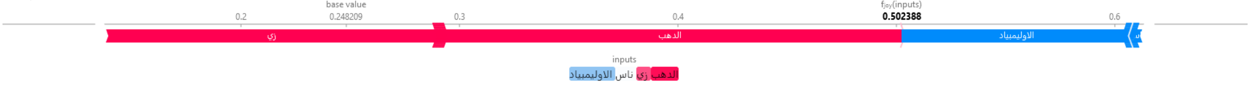


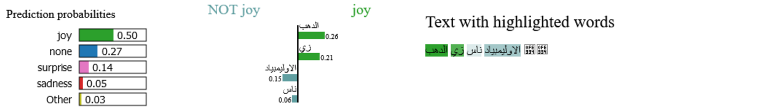
```
index: 1269
True Label none - Predicted Label joy
Explaining according to label joy
العاشره سوداناً اما بعد: الاوليمبياد حقتنا والاولاد اولادنا ولا عزاء للمتشائمين
```





```
index: 1315
True Label none - Predicted Label joy
Explaining according to label joy
□□ .. احنا عددنا ف الاوليمبياد ناس زي الذهب
```



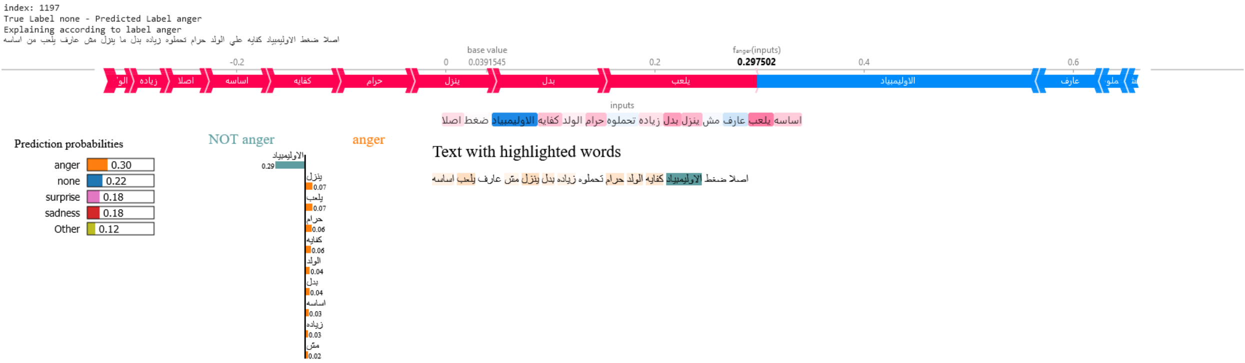


Some examples are understandable why the model predicted them as joy, while others are just confusing to me and the whole context did not help either.

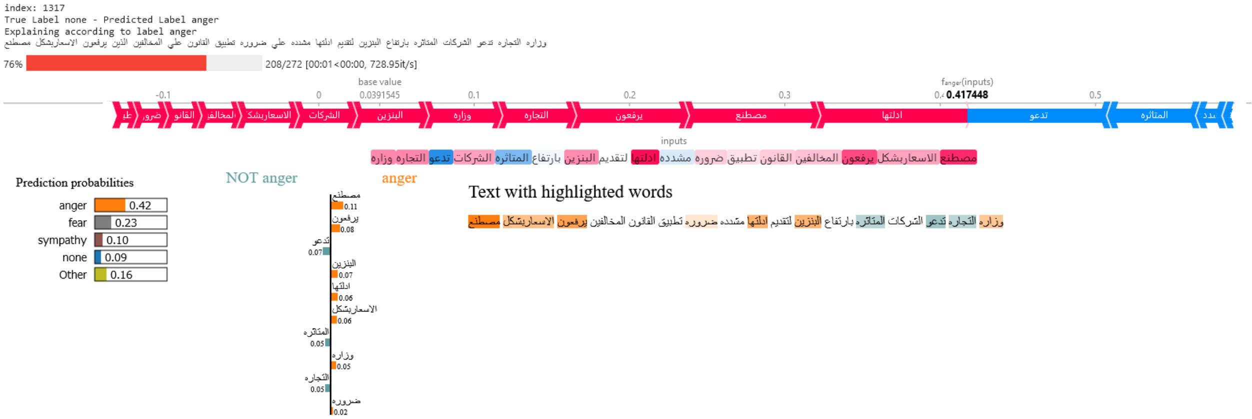
The just context made the model more confused.

So, let's examine something a bit less confusing.

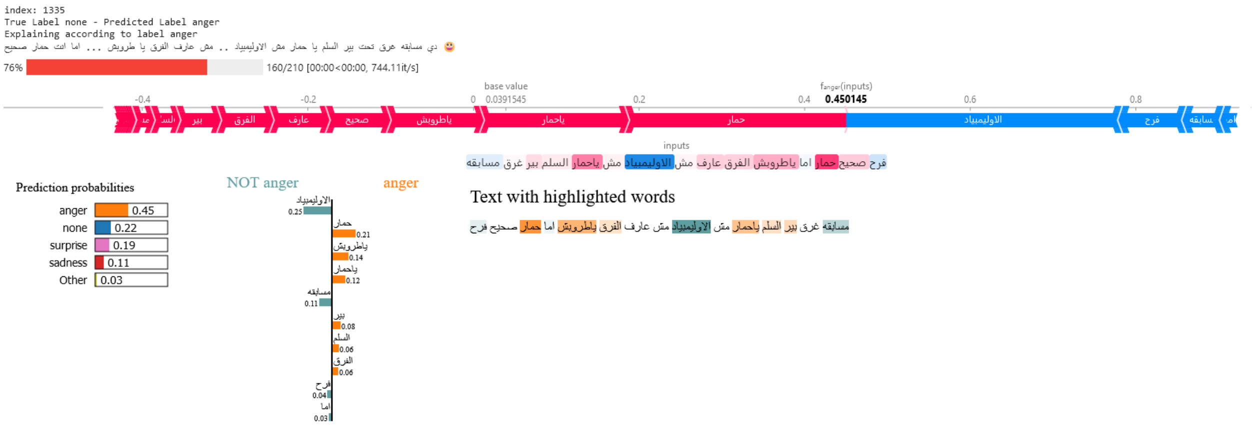
(true = none, pred = anger, total error = 3)



```
index: 1197
True Label none - Predicted Label anger
Explaining according to label anger
اصلا ضغط الاوليمبياد كفايه علي الولد حرام تحملوه زياده بدل ما ينزل مش عارف يلعب من اساسه
```



```
index: 1317
True Label none - Predicted Label anger
Explaining according to label anger
وزاره التجاره تدعو الشركات المتاثره بارتفاع البنزين لتقديم ادلتها مشدده علي ضروره تطبيق القانون علي المخالفين الذين يرفعون الاسعاربشكل مصطنع
76% 208/272 [00:01<00:00, 728.95it/s]
```

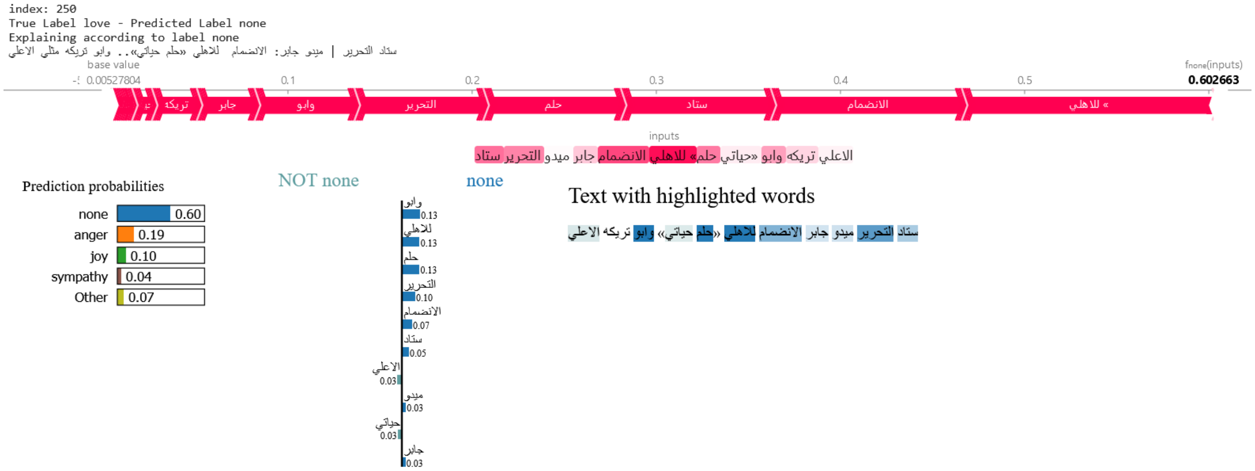


```
index: 1335
True Label none - Predicted Label anger
Explaining according to label anger
دي مسابقه عرق تحت بير السلم يا حمار مش الاوليمبياد .. مش عارف الفرق يا طروبش ... اما انت حمار صحيح 😜
```

76% 160/210 [00:00<00:00, 744.11it/s]

The first and last example are understandable why the model would label them as anger, and I am inclined to agree with its explanation. Especially the last one, as 3 curse words in one sentence is a convincing argument for why it is predicted anger.

(true = love, pred = none, total error = 1)



```
index: 250
True Label love - Predicted Label none
Explaining according to label none
ستاد التحرير | ميدو جابر: الانضمام للاهلي «حلم حياتي».. وابو تريكه مثلي الاعلي
```

Most of the love samples in the dataset are the romantic kind of love, so the problem here is a lack of data in this very specific area.

(true = love, pred = joy, total error = 30)

index: 203

True Label love - Predicted Label joy

Original: احلام العنزي فخر الكويت الف الف مبرووووك بنت قبيلتي وبنت ديرتيييي???????????????????? كفو عيال الديره كل يوم اسم رافع اسمنا بشاره خير ان شاء الله

Cleaned: احلام العنزي فخر الكويت الف الف مبروك بنت قبيلتي وبنت ديرتي كفو عيال الديره يوم اسم رافع اسمنا بشاره خير شاء الله

I would say this one is mixed feelings.

index: 594

True Label love - Predicted Label joy

Original: : الدوله العثمانيه الاسلاميه ستبقي رغم انوف الاعداء ربي احفظ تركيا المسلمه من مكر الماكرين واحفظ كل يلاد المسلمين ♡

Cleaned: الدوله العثمانيه الاسلاميه ستبقي رغم انوف الاعداء ربي احفظ تركيا المسلمه مكر الماكرين واحفظ يلاد المسلمين ♡

This one supports my earlier argument about romantic love and love for one’s country.

index: 823

True Label love - Predicted Label joy

Original: لون حياتك بحب وطاعه الله فهي طويله قصيره ... ايمان تقوه

Cleaned: لون حياتك بحب وطاعه الله طويله قصيره ايمان تقوه

Same as the above.

Conclusion on True = love:

It appears that most of them are predicted wrongly due to a few reasons:

1. The difference between romantic and non-romantic love
2. Mixed feelings
3. Misunderstood context
4. Generally confusing such as the following example:

index: 1430

True Label love - Predicted Label joy

Original: الهلال الساعه 7:30 بالتوفيق شباب الاتحاد &amp الاربعاء ديربي.. اتحاد الشرس

Cleaned: الاربعاء ديربي اتحاد الشرس الهلال الساعه بالتوفيق شباب الاتحاد

(true = surprise, pred = none, total error = 35)

All of them are due to “الاوليمبياد”

ALL OF THEM!!!