Introduction

This report analyzes the prediction explanations produced by LIME library for locally explaining individual predictions.

Note about discretization ranges:

The following attributes are discretized into four categories according to their four quartiles: "age", "mets_achieved", "resting_systolic", "resting_diastolic", "peak_diastolic", "percent_hr_achieved".

Ranges are { [0, 25%), [25%, 50%), [50%, 75%), [75%, 100%)}.

Numbers from 1 to 4 in each attribute maps to one of the following ranges:

- **age**: [18, 40.9, 48.26, 56.23, 79.2]
- **mets_achieved**: [1, 7, 10.1, 12.9, 18]
- **resting_systolic**: [80, 112, 122, 134, 167]
- **resting_diastolic**: [50, 70, 80, 84, 104]
- **peak_diastolic**: [50, 74, 80, 90, 114]
- **percent_hr_achieved**: [0.75, 0.88, 0.93, 0.98, 1.11]

Note about 'reason' attribute:

Number of unique reasons in the data set is 13 reasons. To be able to use SMOTE method for resampling the 'reason' attribute was categorized numerically according to the following mapping:

```
0: 'Ab0rmal Test',
1: 'Arrhythmia',
2: 'Chest Pain',
3: 'Conduction System Disease',
4: 'Dizzy, Fatigue',
5: 'K0wn CAD',
6: 'Other',
7: 'Palpitation',
8: 'Pre-Operation',
9: 'Risk Factor',
10: 'Rule out Ischemia',
11: 'Screening, Research',
12: 'Shortness of Breath'
```

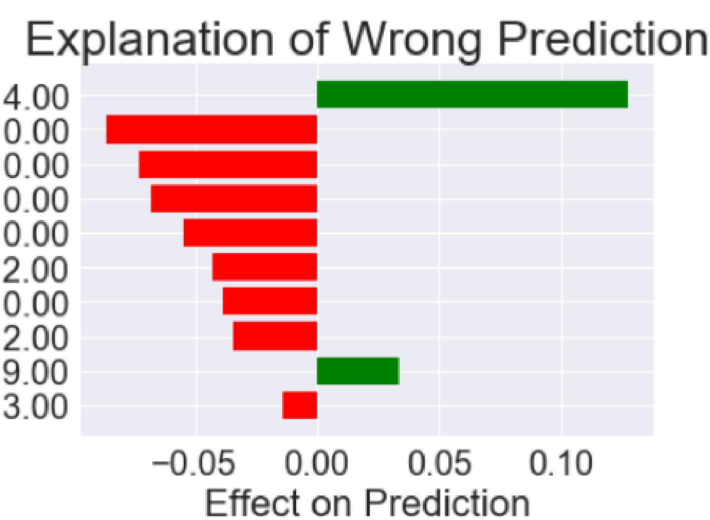
Wrong Predictions Explanation

To get an understanding of what went wrong in **predicting the risk of developing hypertension** of the miss-predicted instances in the test dataset, we will explain two wrongly predicted instances and see what factors forced that prediction.

Instance 1

age	4.0
mets_achieved	2.0
resting_systolic	3.0
resting_diastolic	2.0
peak_diastolic	2.0
reason	10.0
htnresponse	0.0
k0wncad	0.0
dm	0.0
aspirin	0.0
hyperlipid	0.0
black	0.0
percent_hr_achieved	3.0

```
Prediction: 0
Actual Value: 1
```



<u>Analysis:</u>

The plot from LIME shows us how each attribute contributed to the final prediction.

The instance was predicted as (has **no risk** of developing hypertension) while the actual label was the opposite.

There are 8 attributes that forced this prediction:

The most 4 influential attributes that forced the prediction to the wrong way are: k0wncad, htnresponse, dm and black being equal to zero, with **k0wncad** the most influencer towards the wrong prediction.

Other factors forced the wrong prediction with lower effect are:

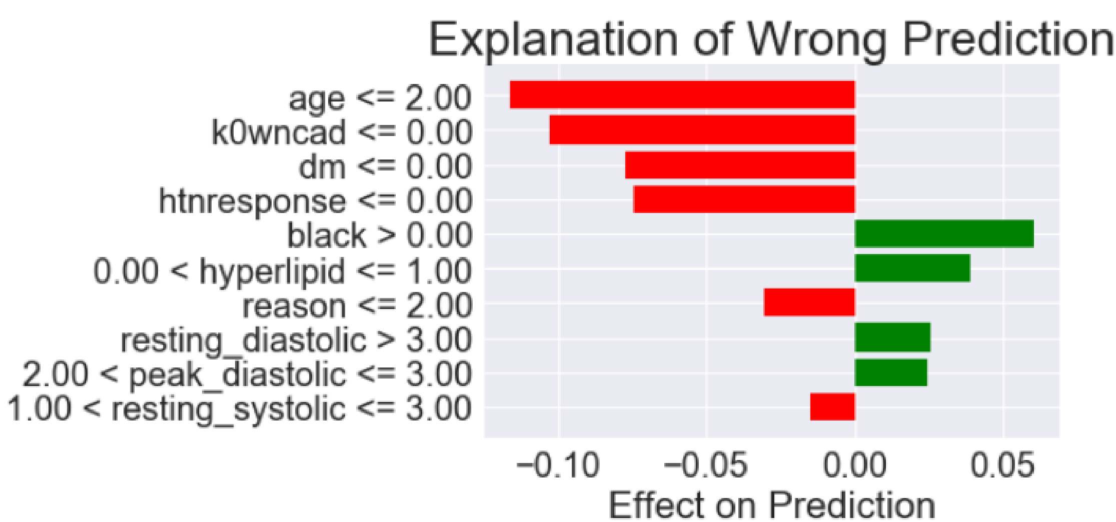
- **resting_diastolic** value being **between 70 and 80**.
- **hyperlibid** being equal to **zero**.
- **peak_diastolic** value being **between 74 and 80**.
- **resting_systolic** falling **between 122 and 134.**

While the previous mentioned attributes decreased the probability of the right prediction there were other two attributes (**age** is being **between 56 and 79** and the **reason** was '**Risk Factor'**) which increased the right prediction probability but their influence weren't strong enough to correctly predict the label.

Instance 2

age	2.0
mets_achieved	1.0
resting_systolic	3.0
resting_diastolic	4.0
peak_diastolic	3.0
reason	2.0
htnresponse	0.0
k0wncad	0.0
dm	0.0
aspirin	0.0
hyperlipid	1.0
black	1.0
percent_hr_achieved	3.0

```
Prediction: 1
Actual Value: 0
```



Analysis:

Remember that the plot from LIME shows us how each attribute contributed to the final prediction.

The instance was predicted (has **risk** of developing hypertension) while the actual label was (not having risk).

There are 4 attributes that affected this prediction the most:

- **black** = 0
- **hyperlibid** =1
- **resting_diastolic** is between **84 and 104.**

- **peak_diastolic** is between **80 and 90**.

Out of the four attributes is known that resting diastolic range actually feels like having risk of hypertension because the normal upper bound number of this measure is 80. The RTF model is trying to make sense of the given attributes but reached the incorrect prediction. However, it is reasonable if the results are evaluated by an expert to see if this kind of wrong prediction is not fatal one.

Other factors decreased this prediction but have less influence:

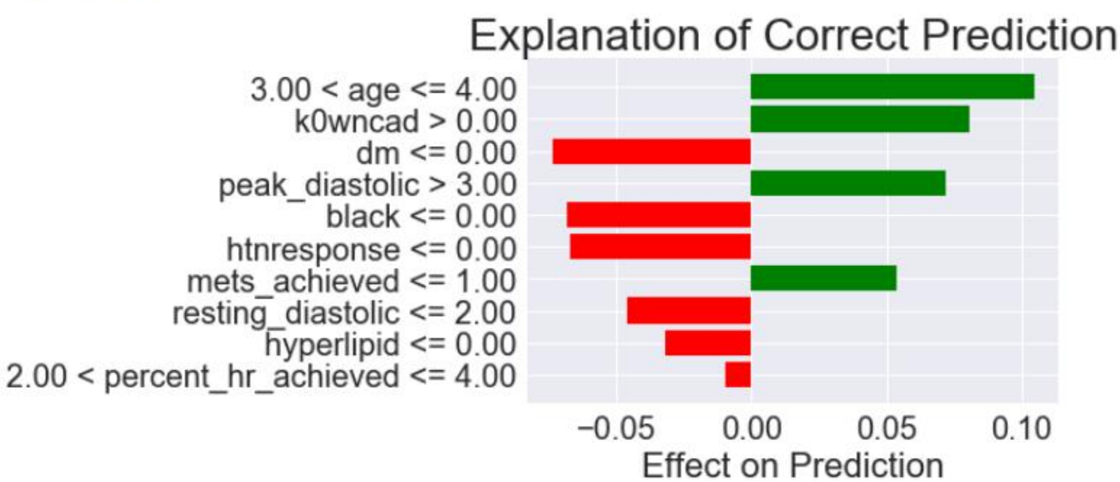
- **age** being **between 48 and 56**.
- **k0wncad** being equal to **zero**.
- **dm** value being **zero**.
- **htnresponse** value being **zero.**

Correct Predictions Explanations

Instance 1

age	4.0
mets_achieved	1.0
resting_systolic	3.0
resting_diastolic	2.0
peak_diastolic	4.0
reason	8.0
htnresponse	0.0
k0wncad	1.0
dm	0.0
aspirin	0.0
hyperlipid	0.0
black	0.0
percent_hr_achieved	4.0

Prediction: 1
Actual Value: 1



<u>Analysis:</u>

The instance was predicted (has **risk** of developing hypertension) correctly as labeled.

There are 4 attributes that affected this prediction the most:

- **age** being **between 56 and 79**.
- **k0wncad** being equal to **One**.
- **peak_diastolic** is between **90 and 114**.
- **Mets_achieved** is between **1 and 7.**

The RTF model is making use of important factors like age and peak diastolic to predict the risk of developing hypertension. From my general knowledge, these two are common factors which affect most human doctors' diagnosis.

Other factors decreased this prediction but have less influence:

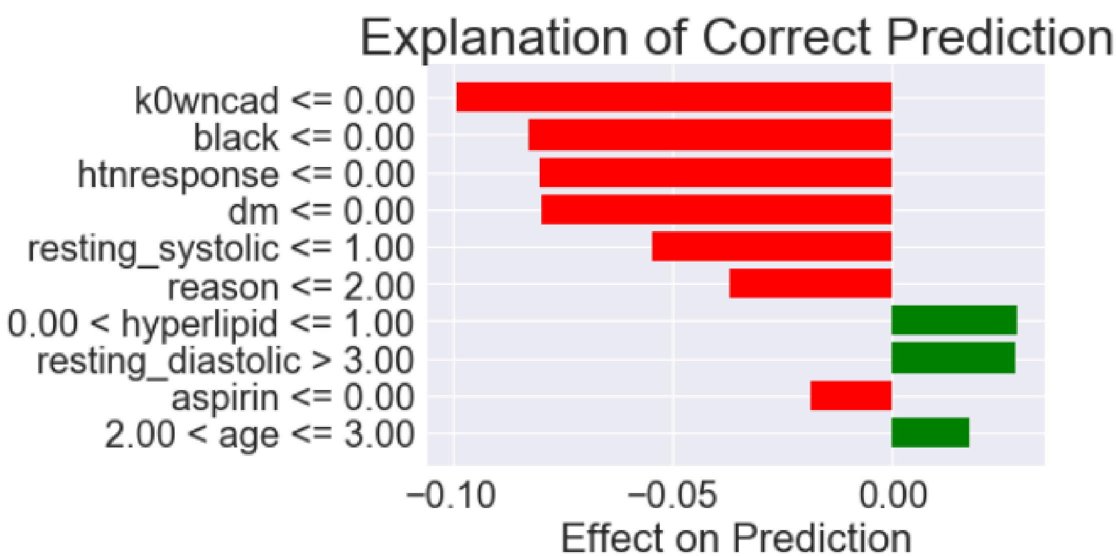
- **dm** value being **zero**.
- **black** value being **zero.**
- **htnresponse** value being **zero.**
- **resting_diastolic** value **between 70 and 80.**
- **hyperlipid** value is **zero.**
- **precent_hr_received** value **between 0.98 and 1.11.**

It seems that the previous 6 attribute weren't enough to force the prediction to the wrong way.

Instance 2

age	3.0
mets_achieved	1.0
resting_systolic	1.0
resting_diastolic	4.0
peak_diastolic	3.0
reason	2.0
htnresponse	0.0
k0wncad	0.0
dm	0.0
aspirin	0.0
hyperlipid	1.0
black	0.0
percent_hr_achieved	3.0

```
Prediction: 0
Actual Value: 0
```



<u>Analysis:</u>

The instance was predicted (has **no risk** of developing hypertension) correctly as labeled.

There are 7 attributes that affected this prediction the most:

- **k0wncad** being equal to **zero**.
- **black** being equal to **zero.**
- **htnresponse** value being **zero.**
- **dm** value being **zero.**

- **resting_systolic** value **between 80 and 112.**
- **reason** being '**Chest Pain'**
- **aspirin** being equal to **zero**.

There were 3 other factors decreased this prediction but had less influence:

- **hyperlipid** value is **one.**
- **resting_diastolic** value **between 80 and 84.**
- **age** is **between 48 and 56.**

It seems that the previous 3 attributes weren't enough to force the prediction to the wrong way.

Conclusion

We have seen the four types of predictions:

- Two wrong predictions:
  - One predicting True to the ML problem while the actual label is False.
  - The other predicts False while the actual label is True.
- Two correct predictions:
  - One predicts True to the ML problem correctly.
  - The other predicts False correctly.

It seems that the False prediction (i.e. predicting the zero label that a patient doesn't have a risk of hypertension) in all the four cases needed more than 7 attributes forcing to that prediction, while predicting True sometimes relied only on 4 attributes. This needs a thorough analysis from an expert answering questions about the quantity vs quality:

- Are the 4 attributes in the True label cases good enough in real world scenarios?
- Can the False label have less than 7 attributes to make that prediction in the previous cases?

It is worth noting that in the wrong predictions there were other factors forcing to the correct label. It is important in this context to ask human experts about both factors. We will call factors forcing to the wrong prediction (which is stronger in this case) S-factors and call the other factors that forced towards the correct miss-labeled prediction as W-factors (S: Strong, W: Weak):

- In a real case scenario, would the S-factors mislead a doctor in the diagnosing? (if yes that means the model is robust compared to the real-world cases)
- In a real case scenario, would the W-factors be sufficient for the miss-labeled prediction?

Other important questions need to be asked to an experienced human in the domain knowledge of the ML problem; because it is important to investigate the robustness of the model in making good predictions and make sure it is not relying on fallacies.