DALEX

Lung cancer case study

Aleksandra Grudziąż
University of Warsaw, MI2 DataLab
Munich, 25.05.2018

Data set

```
  Gender Smoking No.mths.after.giving.up.smoking Industrial.city Chronic.bronchitis Asthma Pack.years.of.smoking Age
1      M     YES                              48              NO                  NO     NO                    40  63
2      M      NO                               0              NO                  NO     NO                     0  69
3      W     YES                               0              NO                 YES     NO                    50  73
4      M     YES                               0              NO                  NO     NO                    35  59
5      M      NO                               0              NO                  NO     NO                    40  70
6      M      NO                               0              NO                 YES     NO                     0  66
  Cancers.in.family Live.more.than.3.y  Lung Degree.of.efficiency.by.Zubord Symptoms.pain Symptoms.haemoptysis
1                NO                  0  Left                               1           YES                   NO
2                NO                  0  Left                               0            NO                   NO
3                NO                  1 Right                               0            NO                   NO
4                NO                  0  Left                               0            NO                   NO
5                NO                  0 Right                               1            NO                   NO
6                NO                  0 Right                               0            NO                   NO
  Symptoms.breathlessness Symptoms.cough Symptoms.degradation.of.efficiency Weight.loss Feature.T Feature.N Feature.M Group
1                     YES            YES                                             NO         0         3         1         0  IIIA
2                      NO             NO                                             NO         0         2         1         0   IIB
3                      NO             NO                                             NO         0         2         0         0    IB
4                      NO             NO                                             NO        30         2         0         0    IB
5                     YES             NO                                             NO         0         2         2         0  IIIA
6                      NO             NO                                             NO         0         3         0         0   IIB
  Initial.histopathological.diagnosis Second.cancer Tumour.volume
1                             Missing            NO     11.300351
2                             Missing            NO     11.831379
3             Non-small cell carcinoma           NO      9.469623
4             Non-small cell carcinoma           NO     10.868568
5                             Missing            NO     12.714395
6                             Missing            NO      9.711116
```

Model

```r
library(mlr)

classif_task <- makeClassifTask(id = "rf_model", data = data_model, target = "Live.more.than.3.y")
classif_lrn_rf <- makeLearner("classif.randomForest", predict.type = "prob")

classif_rf <- train(classif_lrn_rf, classif_task, subset=train_index)
```

Model and its explainer

```r
library(mlr)

classif_task <- makeClassifTask(id = "rf_model", data = data_model, target = "Live.more.than.3.y")
classif_lrn_rf <- makeLearner("classif.randomForest", predict.type = "prob")

classif_rf <- train(classif_lrn_rf, classif_task, subset=train_index)


y_test <- as.numeric(as.character(modelTest$Live.more.than.3.y))

custom_predict_classif <- function(object, newdata) {
        pred <- predict(object, newdata=newdata)
        response <- pred$data[,3]
        return(response)
        }

explainer_classif_rf <- DALEX::explain(classif_rf, data=modelTest, y=y_test,
                                label= "rf", predict_function = custom_predict_classif)
```

Interpretability with DALEX

Model performance

```r
library(DALEX)
mp_classif_rf <- model_performance(explainer_classif_rf)
plot(mp_classif_rf)
```

Model performance

```
library(DALEX)
mp_classif_rf <- model_performance(explainer_classif_rf)
plot(mp_classif_rf)
```
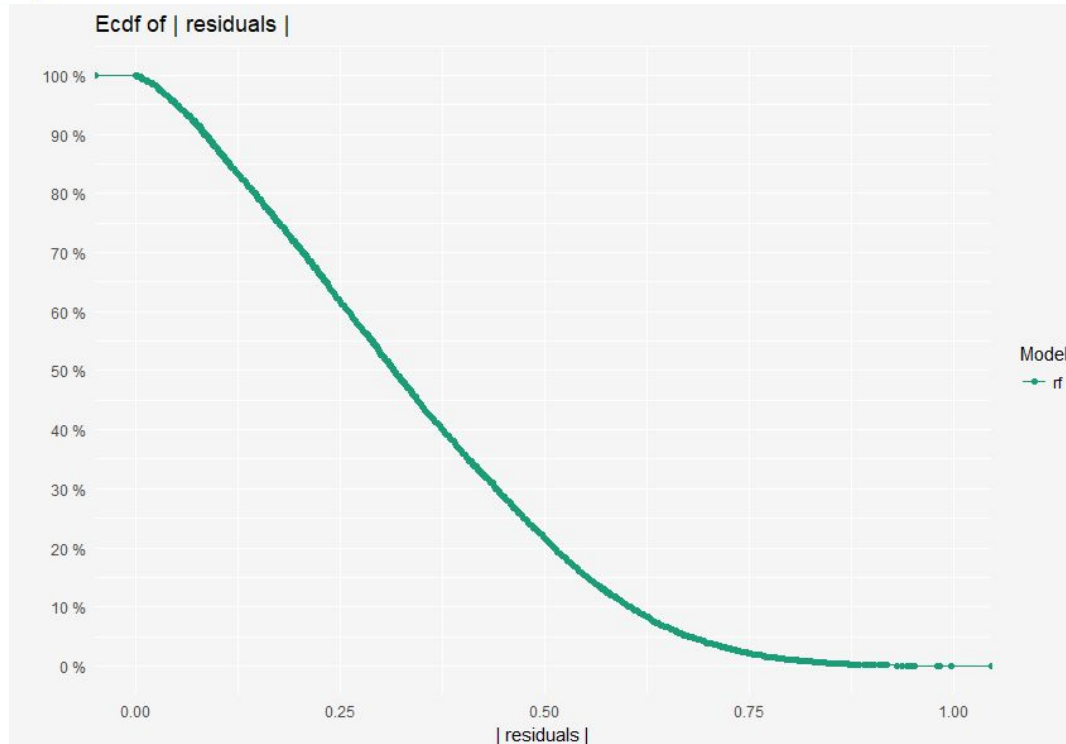
Feature importance

```
vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
vi_classif_rf
```

	variable	dropout_loss	label
1	_baseline_	0.5538452	rf
2	Tumour.volume	0.4593264	rf
3	Age	0.4484891	rf
4	Initial.histopathological.diagnosis	0.4254442	rf
5	Years.of.smoking	0.4241094	rf
6	Group	0.4232901	rf
7	Gender	0.4194648	rf
8	Feature.T	0.4109437	rf
9	Degree.of.efficiency.by.Zubord	0.4086999	rf
10	Lung	0.4063517	rf
11	No.mths.after.giving.up.smoking	0.4049651	rf
12	Feature.N	0.4043632	rf
13	Symptoms.cough	0.4024710	rf
14	Symptoms.degradation.of.efficiency	0.4014890	rf
15	Symptoms.pain	0.4009991	rf
16	Smoking	0.4004038	rf
17	Symptoms.haemoptysis	0.3975389	rf
18	Cancers.in.family	0.3971816	rf
19	Symptoms.breathlessness	0.3969818	rf
20	Industrial.city	0.3964980	rf
21	weight.loss	0.3960647	rf
22	Chronic.bronchitis	0.3937310	rf

Feature importance

```
vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
plot(vi_classif_rf)
```

Feature importance

```
vi_classif_rf <- variable_importance(explainer_classif_rf, loss_function = loss_root_mean_square)
plot(vi_classif_rf)
```
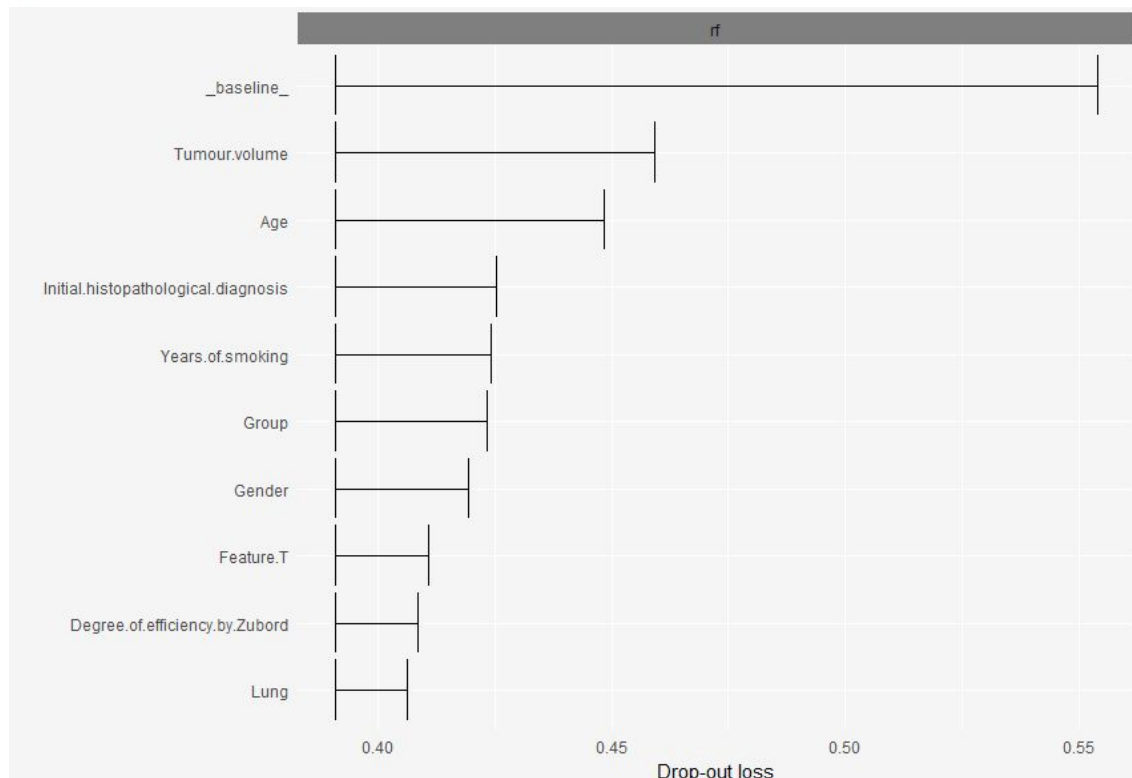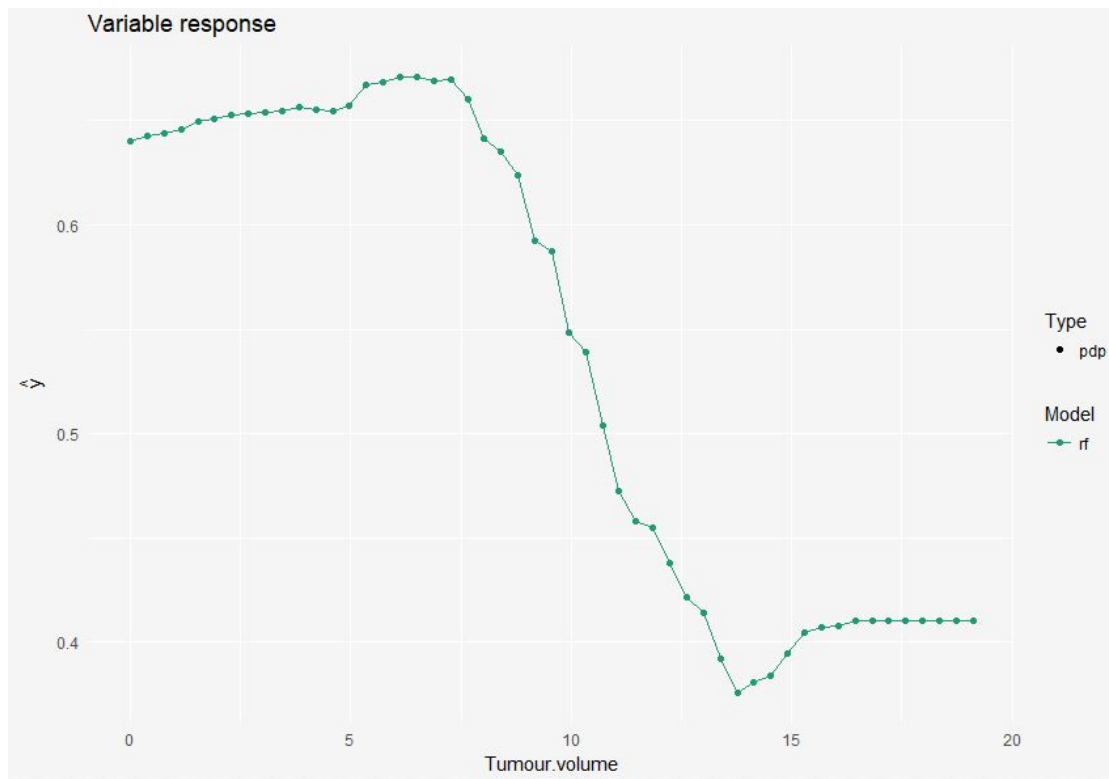
Variable response

```
pdp_classif_rf <- variable_response(explainer_classif_rf, variable = "Tumour.volume", type = "pdp")
plot(pdp_classif_rf)
```
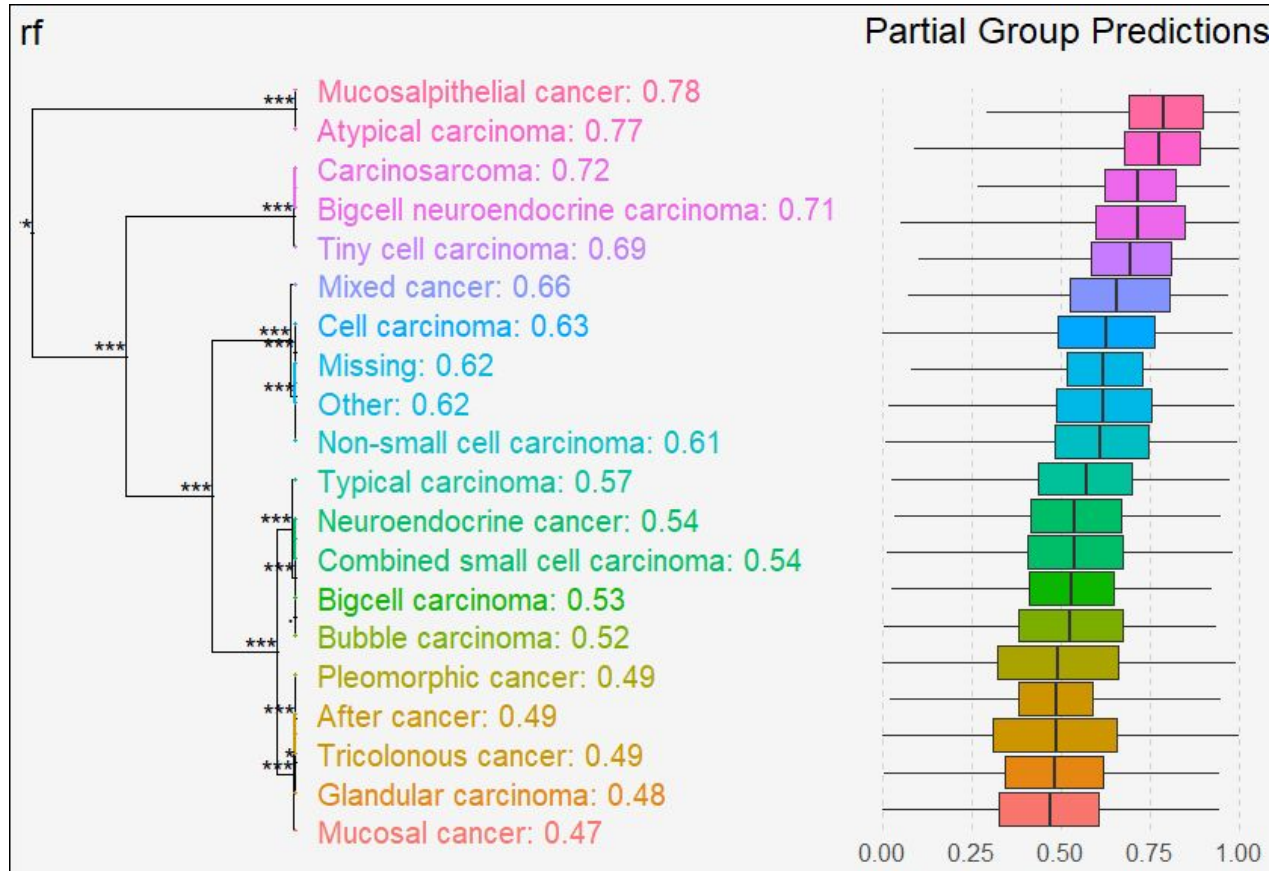
Variable response

```
pdp_classif_rf  <- variable_response(explainer_classif_rf, variable = "Tumour.volume", type = "pdp")
plot(pdp_classif_rf)
```

Merging Path Plot

```r
svd_rf  <- single_variable(explainer_classif_rf,
                           variable = "Initial.histopathological.diagnosis", type = "factor")
plot(svd_rf)
```

Merging Path Plot

breakDown

```
  Gender Smoking No.mths.after.giving.up.smoking Industrial.city Chronic.bronchitis Asthma Years.of.smoking Age
2      M     YES                            1-60              NO                  NO     NO              60  53
  Cancers.in.family  Lung Degree.of.efficiency.by.Zubord Symptoms.pain Symptoms.haemoptysis
2                0 Right                              1            NO                  YES
  Symptoms.breathlessness Symptoms.cough Symptoms.degradation.of.efficiency Weight.loss Feature.T Feature.N
2                     NO             NO                                 NO           0         2         1
  Feature.M Group Initial.histopathological.diagnosis Second.cancer Tumour.volume
2         0   IIB                 Tricolonous cancer            NO      12.07254
```
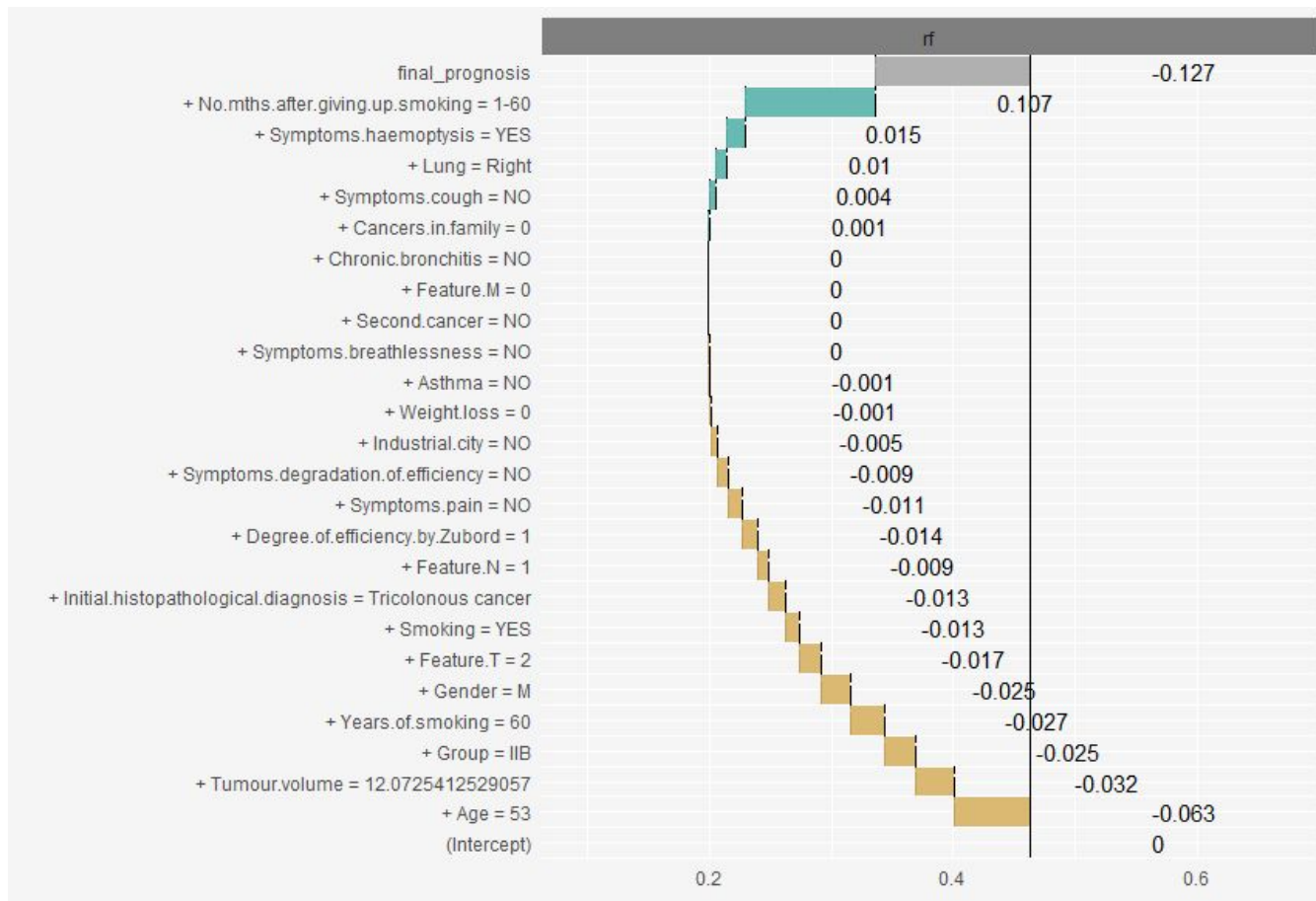
```r
library(breakDown)
library(DALEX)

breakDownExplain <- prediction_breakdown(explainer_classif_rf, modelTest[1,])
plot(breakDownExplain)
```

MI

breakDown

breakDown

```
     Gender Smoking No.mths.after.giving.up.smoking Industrial.city Chronic.bronchitis Asthma Years.of.smoking
44.1     W     YES                              60+             NO                  NO     NO               20
     Age Cancers.in.family Lung Degree.of.efficiency.by.Zubord Symptoms.pain Symptoms.haemoptysis
44.1  55               1 Left                               0            NO                   NO
     Symptoms.breathlessness Symptoms.cough Symptoms.degradation.of.efficiency Weight.loss Feature.T Feature.N
44.1                      NO             NO                                 NO           0         2        0
     Feature.M Group Initial.histopathological.diagnosis Second.cancer Tumour.volume
44.1         0    IB                     Tricolonous cancer            NO      10.79384
```
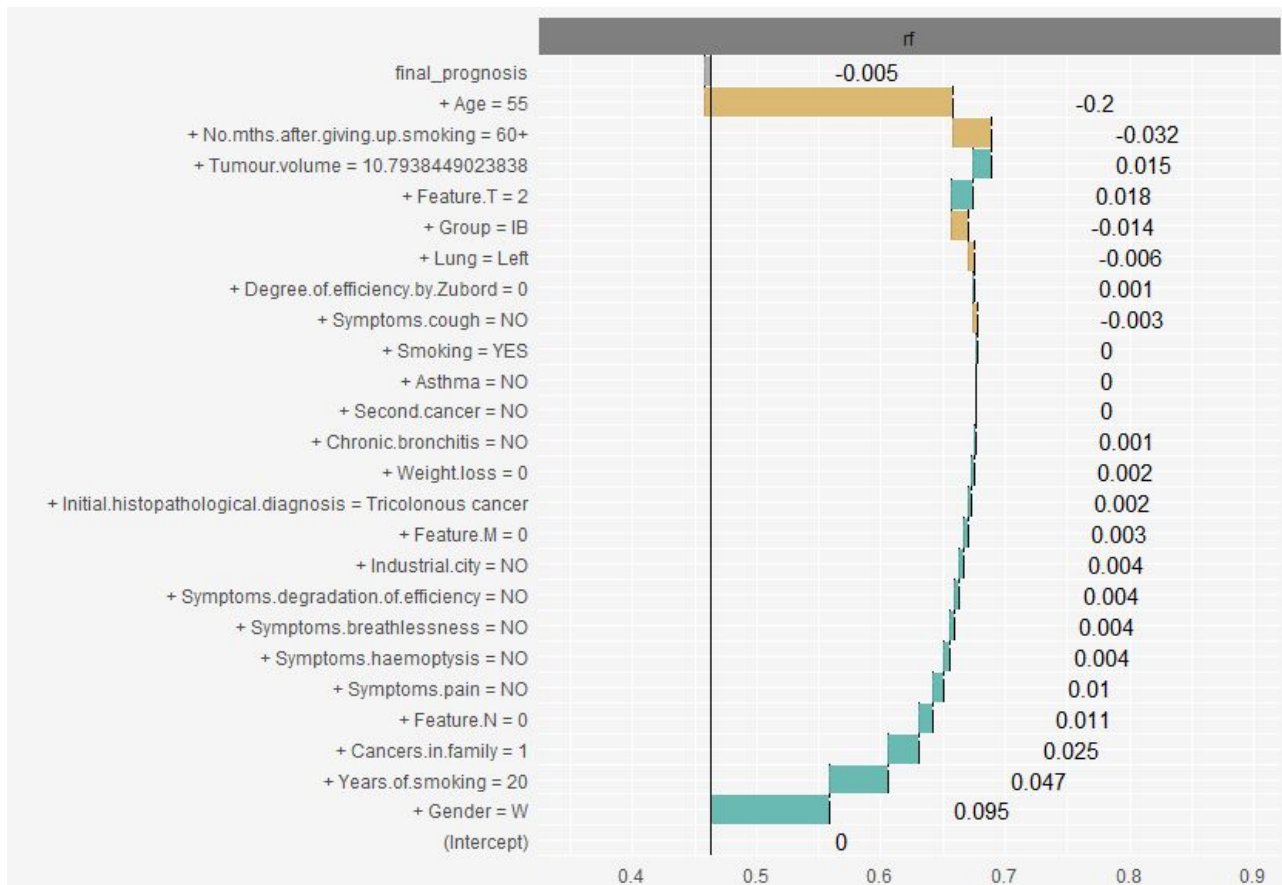
```r
breakDownExplain <- prediction_breakdown(explainer_classif_rf, modelTest[27,])
plot(breakDownExplain)
```

breakDown

Acknowledgements