Résultats Prédiction

# 3 labels

## Case: 3 labels everything

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.173249, Best parameter = {'C': 1.5500000000000005}
* model performance: CV error: mean f1\_macro=0.174 (std=0.030)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.105713, Best parameter = {'C': 0.8000000000000002}
* Parameters kept:

Lasso regression keeps 8 explanatory variables out of 19 and removes 11 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'timeFirstDecay',

'module\_ca400', 'worstProgress\_satisfactory', 'is0Sentiment\_False',

'is0Sentiment\_True']

* Image: OK
* model performance: CV error: mean f1\_macro=0.097 (std=0.017)

Logistic Regression feature selection with RFE

* Parameters kept: The selected 18 features are:

['Supervisor' 'nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet'

'timeFirstDecay' 'module\_ca326' 'module\_ca400' 'module\_ca472'

'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory'

'mainSentiment\_0.0' 'mainSentiment\_1.0' 'mainSentiment\_2.0'

'is0Sentiment\_False' 'is0Sentiment\_True']

* model performance: CV error: mean f1\_macro=0.251 (std=0.051)

Neural Network without feature selection

* GridSearch parameters: 'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05), 'learning\_rate': ['constant','adaptive']

* best GridSearch parameters: Best f1\_macro score = 0.135925, Best parameters = {'alpha': 0.051000000000000004, 'hidden\_layer\_sizes': (16, 6), 'learning\_rate': 'constant'}
* model performance: CV error: mean f1\_macro=0.091 (std=0.052)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method:

['nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet' 'timeFirstDecay'

'module\_ca472' 'module\_mcm' 'module\_pnu' 'mainSentiment\_0.0'

'mainSentiment\_1.0' 'mainSentiment\_2.0']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.235865, Best parameter = {'alpha': 0.451, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'constant'}

* model performance: CV error: mean f1\_macro=0.245 (std=0.045)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 3 8 3

H3 5 42 17

H4 4 38 8

H5 2 31 26

N 0 4 9

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H2 H3 H4 H5 N

trueLabel

H1 0 0 0 1 0

H2 0 4 6 4 0

H3 1 26 15 21 1

H4 0 16 21 10 3

H5 0 11 20 24 4

N 0 2 0 9 2

-------------------------------

Neural Network without feature selection

-------------------------------

nn H2 N

trueLabel

H1 1 0

H2 7 7

H3 30 34

H4 18 32

H5 25 34

N 7 6

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 0 1

H2 3 10 1

H3 11 43 10

H4 12 34 4

H5 9 36 14

N 1 8 4

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.1650459523714846

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.22444621185342287

-------------------------------

Neural Network without feature selection

-------------------------------

0.03874883286647992

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.14946097942533518

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.34328358208955223

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.36318407960199006

-------------------------------

Neural Network without feature selection

-------------------------------

0.06467661691542288

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.2935323383084577

## Case: 3 labels everything without sentiment analysis

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(.5,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.170278, Best parameter = {'C': 1.0000000000000004}
* model performance: CV error: mean f1\_macro=0.181 (std=0.036)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.106159, Best parameter = {'C': 1.2000000000000004}
* Parameters kept:

Lasso regression keeps 8 explanatory variables out of 14 and removes 6 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'timeFirstDecay',

'module\_ca326', 'module\_ca400', 'worstProgress\_moderate',

'worstProgress\_satisfactory']

* Image: OK
* model performance: CV error: mean f1\_macro=0.097 (std=0.017)

Logistic Regression feature selection with RFE

* Parameters kept:

The selected 10 features are:

['nbMeetings' 'timeFirstDecay' 'module\_ca326' 'module\_ca400'

'module\_ca472' 'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory']

* model performance: CV error: mean f1\_macro=0.249 (std=0.047)

Neural Network without feature selection

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters: Best f1\_macro score = 0.136294, Best parameters = {'alpha': 0.201, 'hidden\_layer\_sizes': (16, 6), 'learning\_rate': 'constant'}
* model performance: CV error: mean f1\_macro=0.098 (std=0.046)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method:

['nbMeetings' 'meanTimeMeet' 'module\_ca326' 'module\_ca400' 'module\_ca472'

'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.244309, Best parameter = {'alpha': 0.251, 'hidden\_layer\_sizes': (16, 10), 'learning\_rate': 'adaptive'}

* model performance: CV error: mean f1\_macro=0.252 (std=0.045)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 3 8 3

H3 11 35 18

H4 12 31 7

H5 9 27 23

N 0 8 5

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 6 4 0

H3 23 21 18 2

H4 12 24 10 4

H5 4 27 25 3

N 3 0 6 4

-------------------------------

Neural Network without feature selection

-------------------------------

nn H3 H4 H5

trueLabel

H1 0 0 1

H2 10 0 4

H3 25 3 36

H4 23 1 26

H5 25 2 32

N 6 1 6

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 0 1

H2 0 10 4

H3 1 48 15

H4 0 37 13

H5 2 27 30

N 0 3 10

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.16811850887973204

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.25122969848579607

-------------------------------

Neural Network without feature selection

-------------------------------

0.1253548345596975

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.15120889061187567

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.32338308457711445

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3781094527363184

-------------------------------

Neural Network without feature selection

-------------------------------

0.2885572139303483

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.3383084577114428

## Case: 3 labels everything without satisfaction

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(.5,2,.05)
* best GridSearch parameters Best f1\_macro score = 0.167196, Best parameter = {'C': 0.9500000000000004}
* model performance: CV error: mean f1\_macro=0.173 (std=0.036)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(.1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.105713, Best parameter = {'C': 0.5500000000000002}
* Parameters kept:

Lasso regression keeps 4 explanatory variables out of 15 and removes 11 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'module\_ca400'],

* Image: OK
* model performance: CV error: mean f1\_macro=0.096 (std=0.015)

Logistic Regression feature selection with RFE

* Parameters kept:

The selected 11 features are:

['nbMeetings' 'module\_ca326' 'module\_ca400' 'module\_ca472' 'module\_mcm'

'module\_pnu' 'mainSentiment\_0.0' 'mainSentiment\_1.0' 'mainSentiment\_2.0'

'is0Sentiment\_False' 'is0Sentiment\_True']

* model performance: CV error: mean f1\_macro=0.247 (std=0.051)

Neural Network without feature selection

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters: Best f1\_macro score = 0.135412, Best parameters = {'alpha': 0.35100000000000003, 'hidden\_layer\_sizes': (16, 6), 'learning\_rate': 'adaptive'}
* model performance: CV error: mean f1\_macro=0.101 (std=0.037)

Neural Network feature selection with SFS

* Parameters kept:
* Features selected by the SFS method:

['nbMeetings' 'module\_ca326' 'module\_ca472' 'module\_mcm' 'module\_pnu'

'mainSentiment\_0.0' 'mainSentiment\_1.0' 'mainSentiment\_2.0'

'is0Sentiment\_False' 'is0Sentiment\_True']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.244886, Best parameter = {'alpha': 0.201, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'adaptive'}

* model performance: CV error: mean f1\_macro=0.253 (std=0.050)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 1 10 3

H3 8 25 31

H4 12 25 13

H5 9 21 29

N 1 4 8

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 6 4 0

H3 28 15 20 1

H4 17 17 13 3

H5 11 20 26 2

N 2 0 10 1

-------------------------------

Neural Network without feature selection

-------------------------------

nn H3 H4 N

trueLabel

H1 0 0 1

H2 0 11 3

H3 1 46 17

H4 0 33 17

H5 1 46 12

N 0 10 3

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H4 H5

trueLabel

H1 0 1

H2 9 5

H3 33 31

H4 34 16

H5 27 32

N 4 9

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.15692820012995454

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.20837278381138033

-------------------------------

Neural Network without feature selection

-------------------------------

0.07632446918161205

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.14190361211717525

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.30845771144278605

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3582089552238806

-------------------------------

Neural Network without feature selection

-------------------------------

0.18407960199004975

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.3283582089552239

## Case: 3 labels everything without satisfaction and sentiment analysis

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(.5,2,.05)
* best GridSearch parameters Best f1\_macro score = 0.164743, Best parameter = {'C': 1.2500000000000007}
* model performance: CV error: mean f1\_macro=0.176 (std=0.032)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(.1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.105265, Best parameter = {'C': 0.6000000000000002}
* Parameters kept:
* Lasso regression keeps 4 explanatory variables out of 10 and removes 6 variables.
* Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'module\_ca400']
* Image: OK
* model performance: CV error: mean f1\_macro=0.096 (std=0.015)

Logistic Regression feature selection with RFE

* Parameters kept:

The selected 9 features are:

['Supervisor' 'nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet'

'module\_ca326' 'module\_ca400' 'module\_ca472' 'module\_mcm' 'module\_pnu']

* model performance: CV error: mean f1\_macro=0.208 (std=0.031)

Neural Network without feature selection

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters: Best f1\_macro score = 0.121929, Best parameters = {'alpha': 0.251, 'hidden\_layer\_sizes': (16, 6), 'learning\_rate': 'constant'}
* model performance: CV error: mean f1\_macro=0.112 (std=0.035)

Neural Network feature selection with SFS (STOP)

* Parameters kept:
* Features selected by the SFS method:

['nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet' 'module\_ca326'

'module\_mcm' 'module\_pnu']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.214729, Best parameter = {'alpha': 0.501, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'constant'}

* model performance CV error: mean f1\_macro=0.216 (std=0.038)
* Contigency Tables
* -------------------------------
* Multinomial Logistic Regression without feature selection
* -------------------------------
* mlr H3 H4 H5
* trueLabel
* H1 0 0 1
* H2 3 8 3
* H3 11 23 30
* H4 6 34 10
* H5 7 25 27
* N 1 5 7
* -------------------------------
* Multinomial Logistic Regression feature selection with Lasso
* -------------------------------
* mlr\_lasso H4
* trueLabel
* H1 1
* H2 14
* H3 64
* H4 50
* H5 59
* N 13
* -------------------------------
* Multinomial Logistic Regression feature selection with RFE
* -------------------------------
* mlr\_rfe H3 H4 H5 N
* trueLabel
* H1 0 0 1 0
* H2 4 6 4 0
* H3 20 17 27 0
* H4 17 19 13 1
* H5 8 22 28 1
* N 2 0 10 1
* -------------------------------
* Neural Network without feature selection
* -------------------------------
* nn H3
* trueLabel
* H1 1
* H2 14
* H3 64
* H4 50
* H5 59
* N 13
* -------------------------------
* Neural Network feature selection with SFS
* -------------------------------
* nn\_sfs H4 H5
* trueLabel
* H1 0 1
* H2 4 10
* H3 19 45
* H4 20 30
* H5 12 47
* N 1 12

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.1837094226609323

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.2000876029121589

-------------------------------

Neural Network without feature selection

-------------------------------

0.08050314465408806

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.13969046738192134

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.3582089552238806

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3383084577114428

-------------------------------

Neural Network without feature selection

-------------------------------

0.31840796019900497

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.3333333333333333

# 2 labels

## Case: 2 labels everything

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.181246, Best parameter = {'C': 1.9500000000000013}
* model performance: CV error: mean f1\_macro=0.175 (std=0.032)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters Best f1\_macro score = 0.105713, Best parameter = {'C': 1.0000000000000004}
* Parameters kept:
* Lasso regression keeps 10 explanatory variables out of 18 and removes 8 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'timeFirstDecay',

'module\_ca326', 'module\_ca400', 'worstProgress\_moderate', 'worstProgress\_satisfactory', 'is0Sentiment\_False', 'is0Sentiment\_True'],

* Image: OK
* model performance: CV error: mean f1\_macro=0.097 (std=0.017)

Logistic Regression feature selection with RFE

* Parameters kept: The selected 14 features are:

['nbMeetings' 'timeFirstDecay' 'module\_ca326' 'module\_ca400'

'module\_ca472' 'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory'

'mainSentiment\_0.0' 'mainSentiment\_1.0' 'is0Sentiment\_False', 'is0Sentiment\_True']

* model performance: CV error: mean f1\_macro=0.240 (std=0.044)

Neural Network without feature selection

* GridSearch parameters: 'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05), 'learning\_rate': ['constant','adaptive']

* best GridSearch parameters: Best f1\_macro score = 0.127844, Best parameters = {'alpha': 0.551, 'hidden\_layer\_sizes': (14, 6), 'learning\_rate': 'adaptive'}
* model performance: CV error: mean f1\_macro=0.107 (std=0.040)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method:

['nbMeetings' 'meanTimeMeet' 'module\_ca326' 'module\_mcm' 'module\_pnu'

'worstProgress\_moderate' 'worstProgress\_satisfactory' 'mainSentiment\_1.0'

'is0Sentiment\_False' 'is0Sentiment\_True']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.255492, Best parameter = {'alpha': 0.401, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'constant'}

* model performance: CV error: mean f1\_macro=0.245 (std=0.049)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 3 8 3

H3 10 27 27

H4 6 32 12

H5 6 24 29

N 0 5 8

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 6 4 0

H3 25 21 16 2

H4 14 24 8 4

H5 7 26 23 3

N 4 0 5 4

-------------------------------

Neural Network without feature selection

-------------------------------

nn H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 1 0

H2 0 14 0

H3 2 56 6

H4 2 38 10

H5 0 47 12

N 0 4 9

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.18005690875958613

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.25098760021584543

-------------------------------

Neural Network without feature selection

-------------------------------

0.06640106241699868

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.11178804855275444

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.35323383084577115

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3781094527363184

-------------------------------

Neural Network without feature selection

-------------------------------

0.24875621890547264

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.25870646766169153

## Case: 2 labels everything without sentiment analysis

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.174742, Best parameter = {'C': 1.650000000000001}
* model performance: CV error: mean f1\_macro=0.176 (std=0.031)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters Best f1\_macro score = 0.105713, Best parameter = {'C': 0.7500000000000002}
* Parameters kept:
* Lasso regression keeps 6 explanatory variables out of 14 and removes 8 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'timeFirstDecay',

'module\_ca400', 'worstProgress\_satisfactory']

* Image: OK
* model performance: CV error: mean f1\_macro=0.097 (std=0.017)

Logistic Regression feature selection with RFE

* Parameters kept: The selected 10 features are:

['nbMeetings' 'timeFirstDecay' 'module\_ca326' 'module\_ca400'

'module\_ca472' 'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory']

* model performance: CV error: mean f1\_macro=0.246 (std=0.044)

Neural Network without feature selection

* GridSearch parameters: 'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05), 'learning\_rate': ['constant','adaptive']

* best GridSearch parameters: Best f1\_macro score = 0.116110, Best parameters = {'alpha': 0.651, 'hidden\_layer\_sizes': (14, 6), 'learning\_rate': 'adaptive'}
* model performance: CV error: mean f1\_macro=0.101 (std=0.031)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method:['nbMeetings' 'meanTimeMeet' 'module\_ca326' 'module\_ca400' 'module\_ca472' 'module\_mcm' 'module\_pnu' 'worstProgress\_moderate'

'worstProgress\_satisfactory' 'worstProgress\_unsatisfactory']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.245838, Best parameter = {'alpha': 0.35100000000000003, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'adaptive'}

* model performance: CV error: mean f1\_macro=0.258 (std=0.041)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 1 11 2

H3 3 40 21

H4 7 37 6

H5 1 37 21

N 1 7 5

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 6 4 0

H3 23 21 18 2

H4 12 24 10 4

H5 4 27 25 3

N 3 0 6 4

-------------------------------

Neural Network without feature selection

-------------------------------

nn H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 1 0

H2 0 14 0

H3 0 61 3

H4 1 45 4

H5 1 52 6

N 0 8 5

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.14162214596997205

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.25122969848579607

-------------------------------

Neural Network without feature selection

-------------------------------

0.06640106241699868

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.0909090909090909

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.3034825870646766

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3781094527363184

-------------------------------

Neural Network without feature selection

-------------------------------

0.24875621890547264

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.2537313432835821

## Case: 2 labels everything without satisfaction

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.166354, Best parameter = {'C': 1.600000000000001}
* model performance: CV error: mean f1\_macro=0.162 (std=0.032)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.105713, Best parameter = {'C': 0.8500000000000002}
* Parameters kept:

Lasso regression keeps 5 explanatory variables out of 14 and removes 9 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'module\_ca400',

'is0Sentiment\_False'],

* Image: OK
* model performance: CV error: mean f1\_macro=0.096 (std=0.015)

Logistic Regression feature selection with RFE

* Parameters kept: The selected 13 features are: ['Supervisor' 'nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet' 'module\_ca326' 'module\_ca400' 'module\_ca472' 'module\_mcm' 'module\_pnu' 'mainSentiment\_0.0' 'mainSentiment\_1.0' is0Sentiment\_False'

'is0Sentiment\_True']

* model performance: CV error: mean f1\_macro=0.255 (std=0.052)

Neural Network without feature selection

* GridSearch parameters: 'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05), 'learning\_rate': ['constant','adaptive']

* best GridSearch parameters: Best f1\_macro score = 0.129489, Best parameters = {'alpha': 0.30100000000000005, 'hidden\_layer\_sizes': (14, 6), 'learning\_rate': 'adaptive'}
* model performance: CV error: mean f1\_macro=0.109 (std=0.045)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method: ['nbMeetings' 'meanTimeBetweenMeet' 'module\_ca400' 'module\_ca472' 'module\_mcm' 'module\_pnu' 'mainSentiment\_0.0' 'mainSentiment\_1.0'

'is0Sentiment\_False' 'is0Sentiment\_True']

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.233934, Best parameter = {'alpha': 0.051000000000000004, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'adaptive'}

* model performance: CV error: mean f1\_macro=0.235 (std=0.046)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 3 8 3

H3 13 21 30

H4 10 28 12

H5 4 27 28

N 1 3 9

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 5 5 0

H3 22 17 23 2

H4 22 15 11 2

H5 13 20 24 2

N 2 0 9 2

-------------------------------

Neural Network without feature selection

-------------------------------

nn H2 H3 H5

trueLabel

H1 0 0 1

H2 4 3 7

H3 13 14 37

H4 15 9 26

H5 8 18 33

N 1 6 6

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 0 1

H2 5 6 3

H3 30 11 23

H4 27 12 11

H5 11 23 25

N 2 2 9

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.17946825529950095

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.19682384646694015

-------------------------------

Neural Network without feature selection

-------------------------------

0.13026685415349384

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.1740172160568668

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.34328358208955223

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.31343283582089554

-------------------------------

Neural Network without feature selection

-------------------------------

0.2537313432835821

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.3333333333333333

## Case: 2 labels everything without satisfaction and sentiment analysis

Logistic Regression without feature selection

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.160876, Best parameter = {'C': 1.2000000000000006}
* model performance: CV error: mean f1\_macro=0.161 (std=0.033)

Logistic Regression feature selection with Lasso

* GridSearch parameters: "C":np.arange(1,2,.05)
* best GridSearch parameters: Best f1\_macro score = 0.105683, Best parameter = {'C': 0.5000000000000001}
* Parameters kept:

Lasso regression keeps 4 explanatory variables out of 10 and removes 6 variables.

Selected variables: ['Year', 'Supervisor', 'meanTimeBetweenMeet', 'module\_ca400']

* Image: OK
* model performance: CV error: mean f1\_macro=0.096 (std=0.015)

Logistic Regression feature selection with RFE

* Parameters kept: The selected 9 features are:
* ['Supervisor' 'nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet'
* 'module\_ca326' 'module\_ca400' 'module\_ca472' 'module\_mcm' 'module\_pnu']
* model performance: CV error: mean f1\_macro=0.210 (std=0.033)

Neural Network without feature selection

* GridSearch parameters: 'hidden\_layer\_sizes': [(14,4),(14,6),(14,5),(16,5),(16,6)],

'alpha': np.arange(0.001,1,0.05), 'learning\_rate': ['constant','adaptive']

* best GridSearch parameters: Best f1\_macro score = 0.124714, Best parameters = {'alpha': 0.15100000000000002, 'hidden\_layer\_sizes': (14, 4), 'learning\_rate': 'constant'}
* model performance: CV error: mean f1\_macro=0.081 (std=0.019)

Neural Network feature selection with SFS

* Parameters kept:

Features selected by the SFS method:

['nbMeetings' 'meanTimeBetweenMeet' 'meanTimeMeet' 'module\_ca326'

'module\_ca472' 'module\_pnu'

* GridSearch parameters:

param = {

'hidden\_layer\_sizes': [(14,4),(14,6),(16,6),(16,10),(16,14)],

'alpha': np.arange(0.001,1,0.05),

'learning\_rate': ['constant','adaptive'],

}

* best GridSearch parameters:

Best f1\_macro score = 0.204346, Best parameter = {'alpha': 0.051000000000000004, 'hidden\_layer\_sizes': (16, 14), 'learning\_rate': 'constant'}

* model performance: CV error: mean f1\_macro=0.218 (std=0.043)

Contigency Tables

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

mlr H3 H4 H5

trueLabel

H1 0 0 1

H2 1 10 3

H3 6 27 31

H4 10 27 13

H5 8 22 29

N 1 5 7

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

mlr\_lasso H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

mlr\_rfe H3 H4 H5 N

trueLabel

H1 0 0 1 0

H2 4 6 4 0

H3 20 17 26 1

H4 17 20 12 1

H5 8 22 27 2

N 2 0 10 1

-------------------------------

Neural Network without feature selection

-------------------------------

nn H4

trueLabel

H1 1

H2 14

H3 64

H4 50

H5 59

N 13

-------------------------------

Neural Network feature selection with SFS

-------------------------------

nn\_sfs H3 H4 H5

trueLabel

H1 0 0 1

H2 4 5 5

H3 19 17 28

H4 18 12 20

H5 14 12 33

N 2 1 10

F1-Scores

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.15365107705533235

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.06640106241699868

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.19920874894287469

-------------------------------

Neural Network without feature selection

-------------------------------

0.06640106241699868

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.1640915317110256

Accuracy

-------------------------------

Multinomial Logistic Regression without feature selection

-------------------------------

0.30845771144278605

-------------------------------

Multinomial Logistic Regression feature selection with Lasso

-------------------------------

0.24875621890547264

-------------------------------

Multinomial Logistic Regression feature selection with RFE

-------------------------------

0.3383084577114428

-------------------------------

Neural Network without feature selection

-------------------------------

0.24875621890547264

-------------------------------

Neural Network feature selection with SFS

-------------------------------

0.31840796019900497