

# Analisis de Variables predictoras - sin ciclocursada

Sebastian Jaremczuk

2020-04-17

## carga de datos

```
##
## Recursive feature selection
##
## Outer resampling method: Bootstrapped (10 reps)
##
## Resampling performance over subset size:
##
## Variables Accuracy Kappa AccuracySD KappaSD Selected
##      2      0.6614 0.3014      0.028854 0.06370
##      3      0.6921 0.3699      0.038000 0.08249
##      4      0.7054 0.4015      0.024047 0.04805
##      5      0.7360 0.4627      0.023890 0.04902
##      6      0.7510 0.4942      0.015011 0.02899
##      7      0.7569 0.5066      0.014563 0.02938
##      8      0.7567 0.5068      0.012831 0.02662
##      9      0.7596 0.5126      0.013942 0.02829
##     10      0.7624 0.5182      0.010534 0.02149
##     11      0.7629 0.5195      0.008171 0.01640
##     12      0.7664 0.5264      0.011269 0.02214
##     13      0.7629 0.5191      0.010828 0.02189
##     14      0.7674 0.5280      0.009451 0.01906
##     15      0.7720 0.5368      0.013492 0.02714
##     16      0.7710 0.5348      0.014071 0.02794
##     17      0.7711 0.5347      0.016302 0.03267
##     18      0.7733 0.5390      0.016023 0.03303
##     19      0.7708 0.5340      0.015181 0.03001
##     20      0.7737 0.5400      0.014674 0.02969
##     21      0.7720 0.5363      0.014770 0.02990
##     22      0.7715 0.5355      0.016211 0.03274
##     23      0.7748 0.5422      0.015507 0.03147      *
##
## The top 5 variables (out of 23):
##      tipo_de_aprobacion_libre , Turno_Tarde , Turno_Noche , tipo_de_aprobacion_no_firmo , Apr

## [1] "tipo_de_aprobacion_libre"      "Turno_Tarde"
## [3] "Turno_Noche"                  "tipo_de_aprobacion_no_firmo"
## [5] "Aprobado"                      "Turno_Manana"
## [7] "tipo_de_aprobacion_firmo"      "Nota_max_prom"
## [9] "cant_recursada_regular_No_Recurso" "cant_resursada_regular"
```

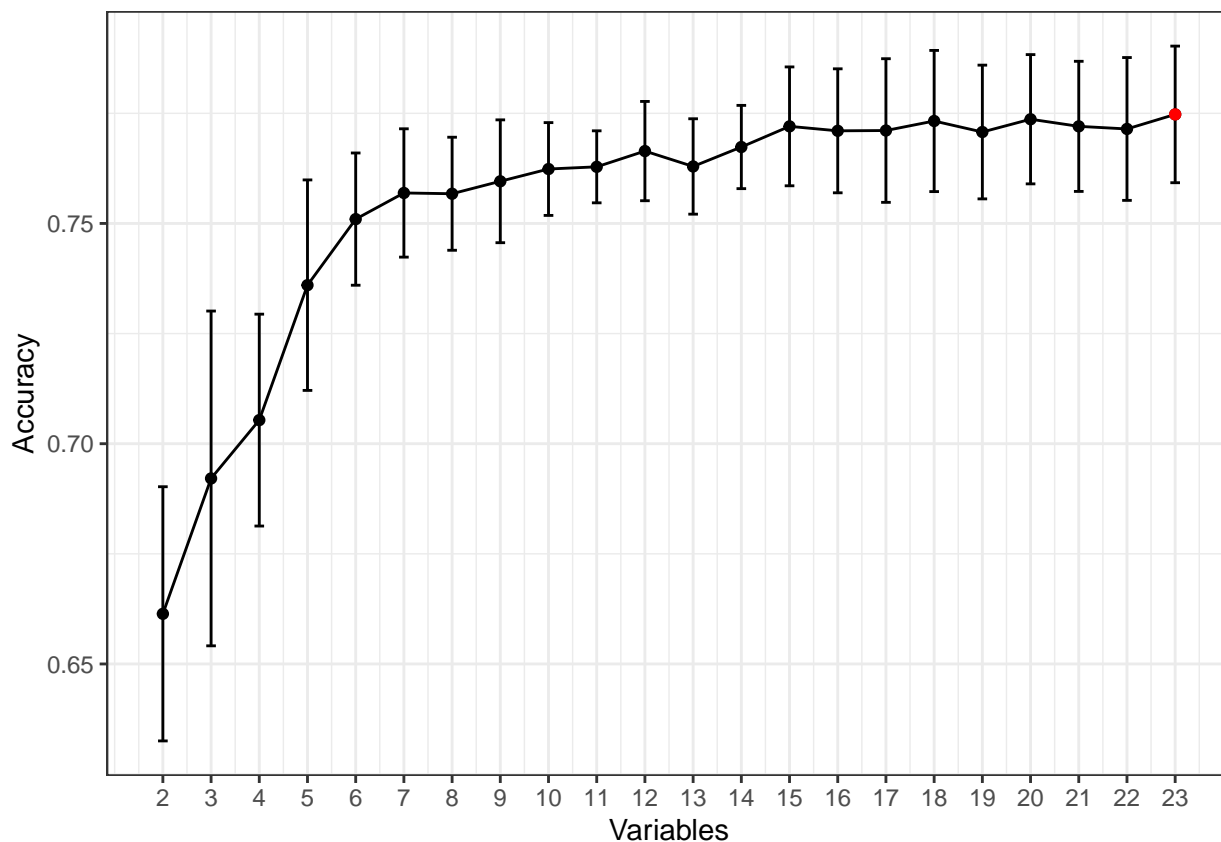
```

## [11] "edad_al_ingreso " "cant_rekursada_regular_Recurso1vez "
## [13] "noAprobado " "tipo_de_aprobacion_cambio_curso "
## [15] "Promociono " "tipo_de_aprobacion_promociono "
## [17] "Nota " "EsTecnico_X1 "
## [19] "cant_rekursada_regular_Recurso2vez " "Sexo_M"
## [21] "cant_rekursada_regular_Recurso3vez " "EsTecnico_SinDato "
## [23] "cant_rekursada_regular_Recurso4vez "

```

Table 1: Top 10 Modelos con cantidad de variables seleccionadas según Accuracy

Variables	media_accuracy	media_kappa
23	0.7747542	0.5421768
20	0.7736578	0.5399892
18	0.7732554	0.5390105
15	0.7720421	0.5368054
21	0.7720411	0.5363012
22	0.7714524	0.5355353
17	0.7710945	0.5346718
16	0.7710278	0.5348438
19	0.7707567	0.5340365
14	0.7673518	0.5279523



El mejor resultado se consigue con modelos que tienen las 23 variables mencionadas anteriormente y en el grafico puede verse la evolución de dicha métrica en función de la cantidad de variables que usa con su dispersión según todos los modelos generados con esa combinación.

Se puede observar que desde que la cantidad de variables es 15, es un valor muy próximo al máximo que se consigue con 23 variables. por lo que si se opta por este tipo de dataset y se lo quisiera explicar quizás sería mas sencillo con menos variables aunque sigue siendo una cantidad considerable y dependerá mucho de la interacción entre ellas.

Tras ajustar cada modelo, se recalcula la influencia de cada variable. De esta forma, para cada tamaño de modelo, se obtiene un ranking de la importancia promedio de las variables.

Table 2: Influencia de variables en el resultado

var	media_influencia	sd_influencia
tipo_de_aprobacion_libre	50.79648	2.5524875
Turno_Tarde	44.59747	0.8908022
Turno_Noche	43.09091	1.1487629
tipo_de_aprobacion_no_firmo	42.65281	1.5211102
Aprobado	42.15115	2.1368377
tipo_de_aprobacion_firmo	40.75425	1.1106366
Turno_Manana	40.69317	1.9193145
Nota_max_prom	40.54109	1.7290803
cant_recurzada_regular_No_Recurso	36.59649	1.7624988
cant_resursada_regular	36.06834	1.6028767
edad_al_ingreso	34.36706	2.0057595
cant_recurzada_regular_Recurso1vez	33.45997	1.0793458
noAprobado	32.05701	1.4956207
tipo_de_aprobacion_cambio_curso	32.04880	1.2241621
tipo_de_aprobacion_promociono	30.37739	0.4212574
Promociono	30.31098	1.9463734
Nota	30.01460	1.9773123
EsTecnico_X1	24.08818	1.1924095
cant_recurzada_regular_Recurso2vez	23.93968	1.2196074
Sexo_M	18.84818	2.2983891
cant_recurzada_regular_Recurso3vez	17.05511	1.4083702
EsTecnico_SinDato	16.70716	1.3843021
cant_recurzada_regular_Recurso4vez	13.33266	1.3074133

