

D2MCS:

Un paquete en R para desenvolver y desplegar automáticamente un Sistema Multi-Clasificador

Autores: M. Ferreiro-Díaz, D. Ruano-Ordás, J. R. Méndez

CRAN: <https://CRAN.R-project.org/package=D2MCS>

GitHub: <https://github.com/drordas/D2MCS>

ÍNDICE

1.Descripción del problema

2.Propuesta

3.Objetivos

4.Arquitectura

5.Modo de funcionamiento

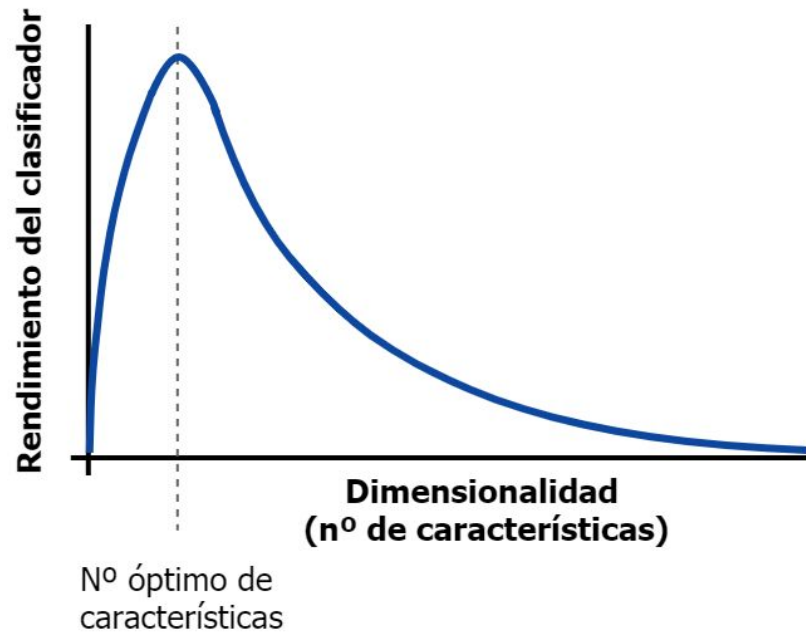
6.Ejemplo

7.Conclusiones

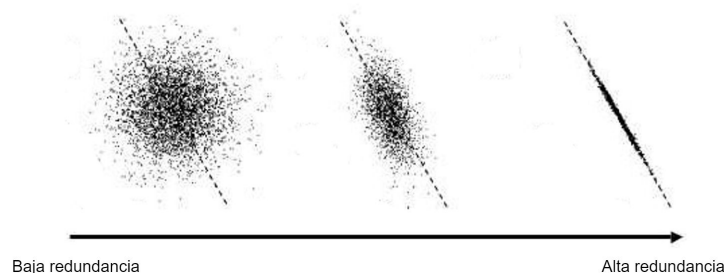
8.Referencias

DESCRIPCIÓN DEL PROBLEMA

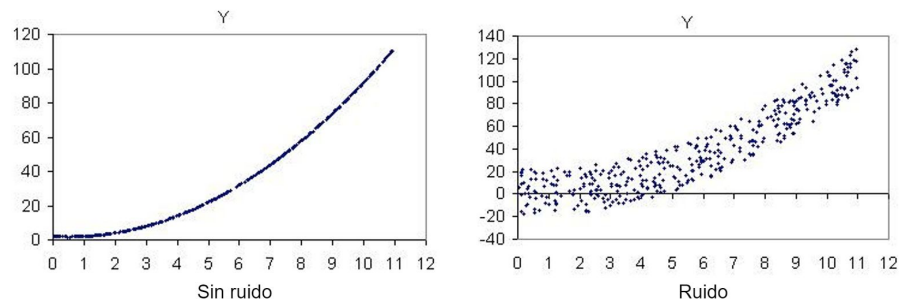
La maldición de la dimensionalidad



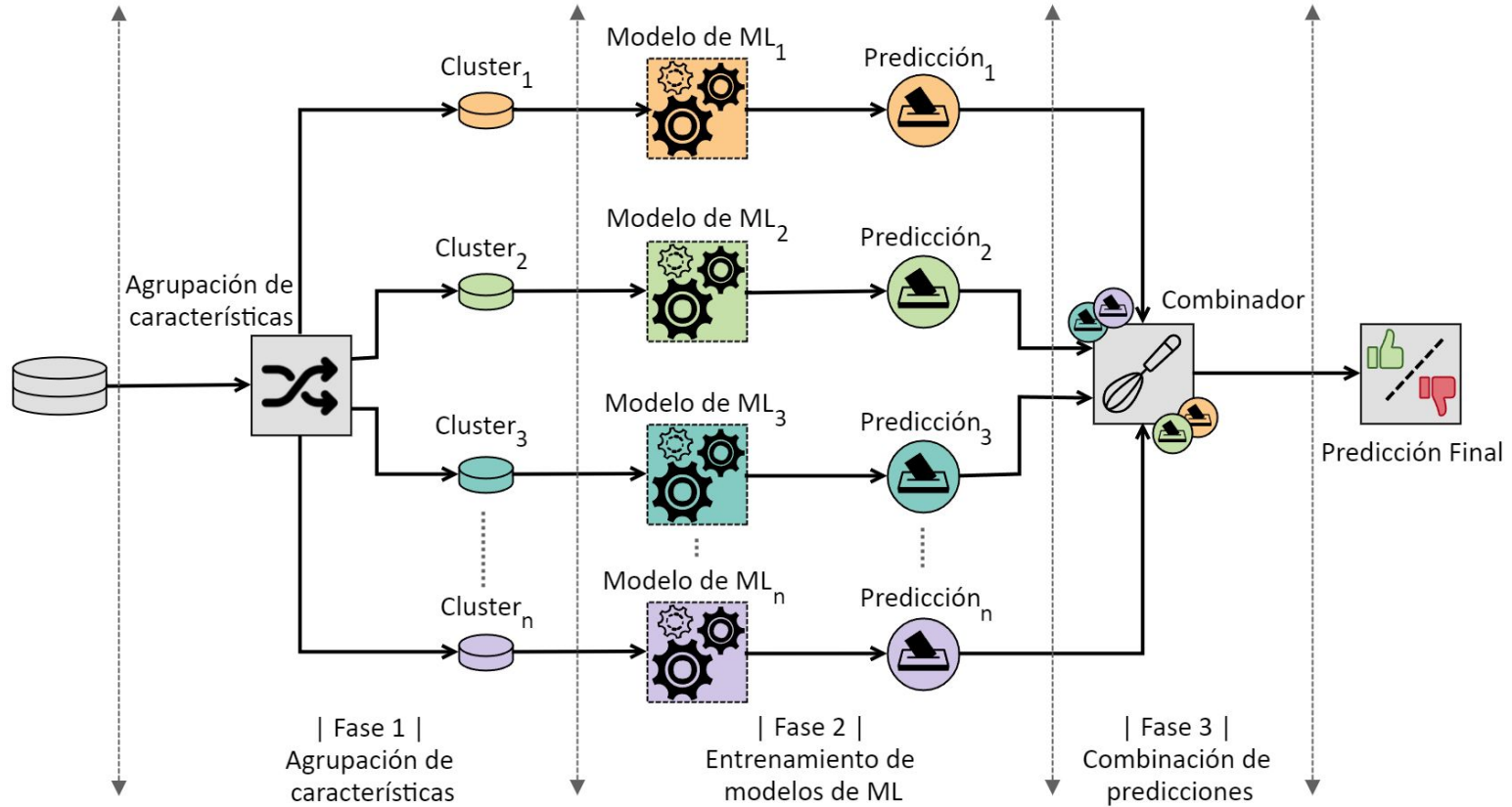
Incrementa la redundancia (información duplicada)



Incrementa el ruido de los datos



PROPUESTA

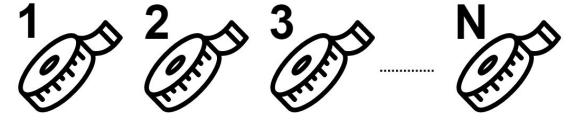


OBJETIVOS

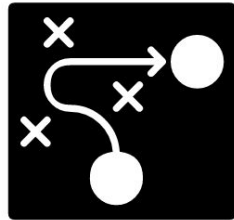
**Manipulación
de datos**



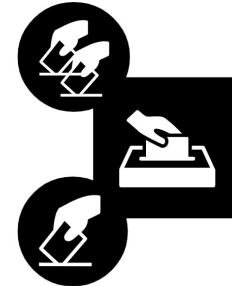
**Medidas
de evaluación**



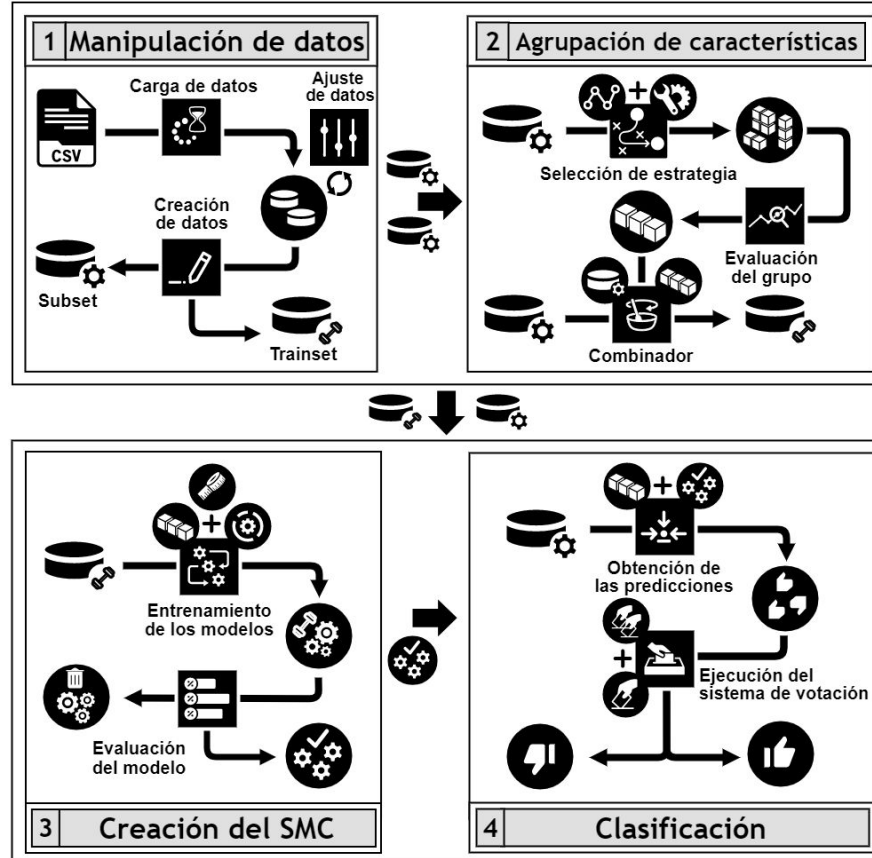
**Estrategias
de clustering**



**Mecanismos
de votación**



ARQUITECTURA



MODO DE FUNCIONAMIENTO (I)

Estrategias de clustering

SimpleStrategy

TypeBasedStrategy

DependencyBasedStrategy

Heurísticas

ChiSquareHeuristic

FisherTestHeuristic

GainRatioHeuristic

InformationGainHeuristic

KendallHeuristic

MCCHeuristic

MultinformationHeuristic

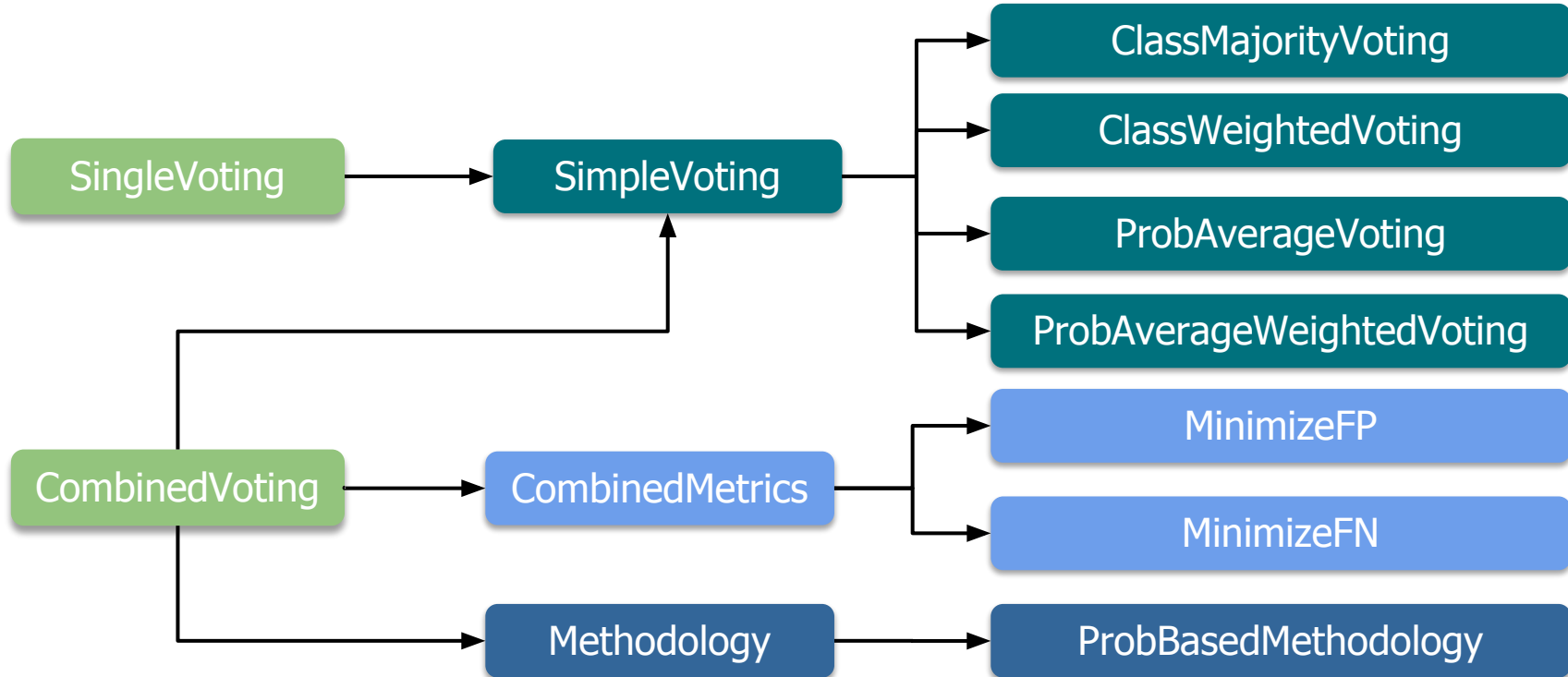
OddsRatioHeuristic

PearsonHeuristic

SpearmanHeuristic

MODO DE FUNCIONAMIENTO (II)

Sistemas de votación



EJEMPLO (I)

Fase 1: Manipulación de datos

Conjunto de datos

Temática:

- Componentes de medicamentos

Dimensiones:

- Filas: 3924
- Columnas: 2133

Columna objetivo:

- *Activity*
 - 1: Compuesto activo
 - 0: Compuesto no activo

```
01 loader <- DatasetLoader$new()
02
03 data <- loader$load(filepath = "medicamentos.csv", header = TRUE,
04                      normalize.names = TRUE, sep = "\\t")
05
06 data$createPartitions(num.folds = 4, class.balance = "Activity")
07
08 cluster.subset <- data$createSubset(num.folds = c(1, 2), class.index = "Activity",
09                                    positive.class = "Active")
10
11 train.subset <- data$createSubset(num.folds = c(2, 3), class.index = "Activity",
12                                  positive.class = "Active")
13
14 test.subset <- data$createSubset(num.folds = 4, class.index = "Activity",
15                                 positive.class = "Active")
```

EJEMPLO (II)

Fase 2: Agrupación de características

```
01 strategy <- DependencyBasedStrategy$new(  
02     subset = cluster.subset,  
03     heuristic = c(MCCHeuristic$new(),  
04                 InformationGainHeuristic$new()),  
05     configuration = DependencyBasedStrategyConfiguration$new())  
06  
07 strategy$execute()  
08  
09 train.set <- strategy$createTrain(subset = train.subset)
```

EJEMPLO (III)

Fase 3: Creación del SMC

```
01 ex.classifiers <- c("naive_bayes", "ranger", "svmRadialWeights")
02
03 trFunction <- TwoClass$new(method = "cv", number = 10, savePredictions = "final",
04                             classProbs = TRUE, allowParallel = TRUE,
05                             verboselter = FALSE, seed = 1234)
06
07 d2mcs <- D2MCS$new(dir.path = "exampleDependencyBased")
08
09 trained.models <- d2mcs$train(train.set = train.set,
10                               train.function = trFunction,
11                               ex.classifiers = ex.classifiers,
12                               metrics = c("MCC", "PPV"))
```

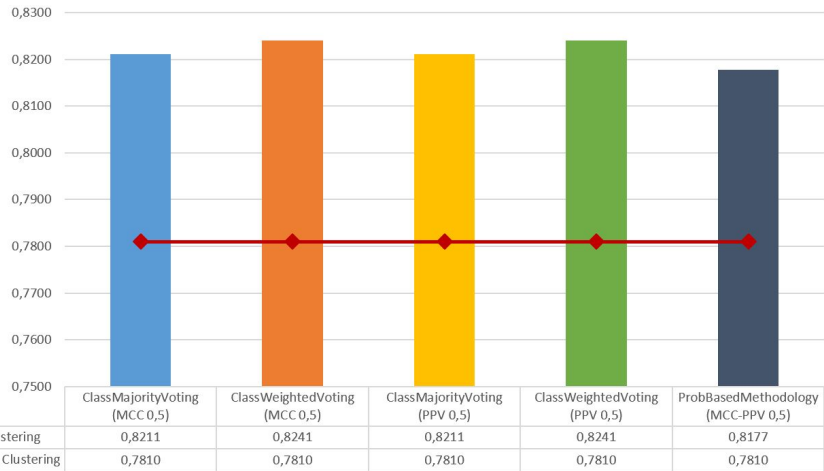
EJEMPLO (IV)

Fase 4: Clasificación

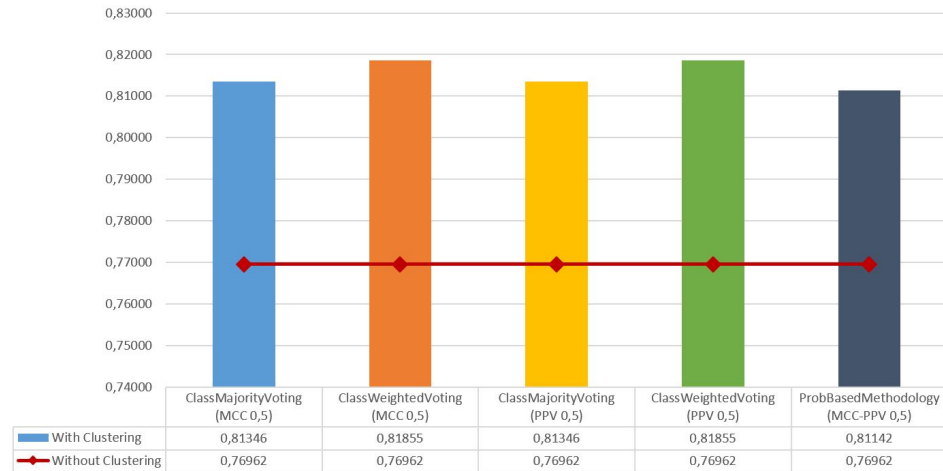
```
01 predictions <- d2mcs$classify(  
02     train.output = trained.models,  
03     subset = test.subset,  
04     voting.types = c(  
05         SingleVoting$new(voting.schemes = c(ClassMajorityVoting$new(),  
06                                             ClassWeightedVoting$new()),  
07         metrics = c("MCC", "PPV")),  
08     CombinedVoting$new(voting.schemes = ClassMajorityVoting$new(),  
09                         combined.metrics = MinimizeFP$new(),  
10                         methodology = ProbBasedMethodology$new(),  
11                         metrics = c("MCC", "PPV"))),  
12     positive.class = "Active")  
13  
14 predictions$getPerformances(test.subset,  
15     measures = list(Accuracy$new(), MCC$new(),  
16                     PPV$new(), TP$new(), TN$new(),  
17                     FP$new(), FN$new()))
```

EJEMPLO (V)

FScore

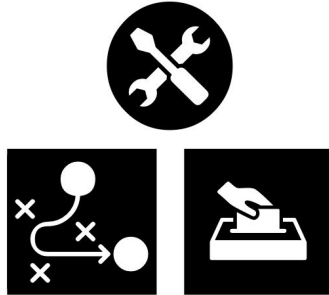


Accuracy

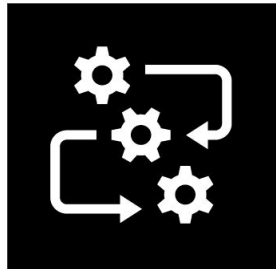


CONCLUSIONES

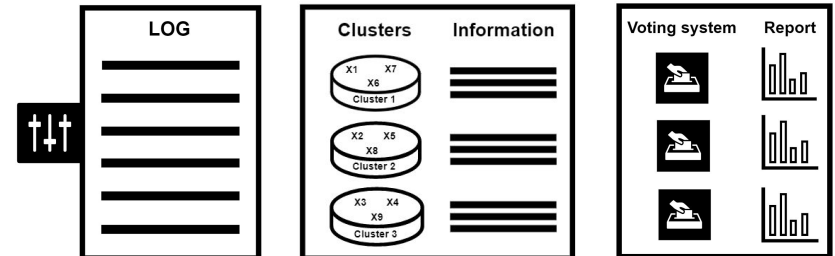
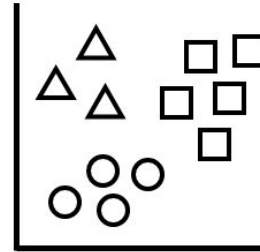
Personalización de funcionalidades



Entrenamiento de modelos de ML



Futuras ampliaciones



REFERENCIAS

- Paquete D2MCS:
 - CRAN: <https://CRAN.R-project.org/package=D2MCS>
 - GitHub: <https://github.com/drordas/D2MCS>
- Artículos relacionados:
 - <https://doi.org/10.1186/s13321-019-0389-9>
 - <https://doi.org/10.1016/j.eswa.2018.12.032>
- Librerías utilizadas:

Suggests		Imports			
grDevices	knitr	caret	devtools	dplyr	FSelector
rmarkdown	testthat	ggplot2	ggrepel	gridExtra	infotheo
		mccr	mltools	ModelMetrics	questionr
		recipes	R6	tictoc	varhandle

D2MCS:

Un paquete en R para desenvolver y desplegar automáticamente un Sistema Multi-Clasificador

Autores: M. Ferreiro-Díaz, D. Ruano-Ordás, J. R. Méndez

CRAN: <https://CRAN.R-project.org/package=D2MCS>

GitHub: <https://github.com/drordas/D2MCS>