

26

DETERMINACIÓN DE RASGOS ÚTILES AL PROCESO DE APRENDIZAJE AUTOMÁTICO SOBRE EL CONJUNTO DE DATOS EN ESTUDIANTES DE 1ER Y 2DO AÑO DE INGENIERÍA INFORMÁTICA EN CIENFUEGOS

DETERMINATION OF THE MAIN ATTRIBUTES OF THE MATCHING LEARNING PROCESS IN DATA SET IN STUDENTS OF INFORMATION TECHNOLOGIES IN CIENFUEGOS

MSc. Viviana R. Toledo Rivero¹

E-mail: paulioski2015@gmail.com

MSc. Carlos Aguilar Aguilar¹

MSc. Kadir Héctor Ortiz¹

E-mail: kadirhectorortiz@gmail.com

¹Universidad Metropolitana del Ecuador. República del Ecuador.

¿Cómo referenciar este artículo?

Toledo Rivero, V.R., Aguilar Aguilar, C., & Ortiz, K.H. (2014). Determinación de rasgos útiles al proceso de aprendizaje automático sobre el conjunto de datos en estudiantes de 1er y 2do año de Ingeniería Informática en Cienfuegos. Revista Conrado [seriada en línea], 12(53), pp. 178-189. Recuperado de <http://conrado.ucf.edu.cu/>

RESUMEN

La carrera de Ingeniería Informática de la Universidad de Cienfuegos enfrenta dificultades con los bajos resultados académicos obtenidos en la carrera, fundamentalmente en los primeros años, que constituyen asignaturas precedentes a las del ciclo especializado. Por tal motivo, en aras de buscar una solución para obtener mejores resultados académicos se propone un modelo de predicción haciendo uso de aprendizaje automático. Como requisito previo se realiza una selección de rasgos. Para este trabajo se ha sido escogido el programa WEKA por las facilidades que brinda para la realización de diferentes pruebas sobre el conjunto de datos. Se emplean dos métodos de evaluación con diferentes métodos de búsqueda, dando como resultados nuevos conjuntos de datos reducidos. Estos conjuntos garantizan el entrenamiento eficiente de los algoritmos a utilizar para la toma de decisiones oportunas en la carrera.

Palabras clave:

Selección de rasgos, atributos, conjuntos de datos.

ABSTRACT

At the University of Cienfuegos the students of Information Technologies have some difficulties to successfully pass the first years of the specialty, mainly because these subjects are precedents to another one in the specialty. The paper is aimed to propose a model of prediction using automatic learning. As a prerequisite a data mining is developed. The WEKA Software is used due to its facilities for testing the data set. Two methods of evaluation with different methods of search are used resulting in new data set with less quantity of attributes. The new groups guarantee the efficient training of algorithms to be used when making decisions in the specialty.

Keywords:

Data mining, attributes, data set.

INTRODUCCIÓN

En el último lustro se ha observado en la carrera de Ingeniería Informática de la UCF que los estudiantes confrontan dificultades académicas, que se traducen posteriormente en un pobre rendimiento, fundamentalmente en el primer y segundo años, lo que con frecuencia trae consigo falta de motivación por los estudios, alumnos que repiten, y en el peor de los casos deserción escolar.

Con el objetivo de explicar este fenómeno, y con la idea de disminuir sus efectos, se decide aplicar técnicas de aprendizaje automático que permitan descubrir las causas principales de tal comportamiento. Previo a estas técnicas se requiere depurar el conjunto de datos, haciendo un análisis para determinar el conjunto de los rasgos más importantes que deberán ser utilizados en el proceso de aprendizaje automático.

Es notable la cantidad de ejemplos donde el desempeño de los algoritmos de clasificación puede ser mejorada si se eliminan atributos irrelevantes. Aunque la mayoría de los algoritmos de aprendizaje automático están diseñados para aprender cuáles son los atributos apropiados para tomar decisiones, en la práctica cuando existen atributos irrelevantes frecuentemente se “confunden”. Debido al efecto negativo que sobre el aprendizaje provocan los mismos -de 5 a un 10 % de deterioro en situaciones probadas sobre árboles de decisión - es común preceder al aprendizaje con la selección de atributos que elimine sino todos, la mayoría de los atributos irrelevantes.

En este artículo explicamos el proceso de selección de atributos que precede a la utilización de técnicas como las redes neuronales y árboles de decisión sobre un conjunto de datos de los estudiantes de 1ro y 2do años de la carrera de Ingeniería Informática en la Universidad de Cienfuegos.

DESARROLLO

La selección de rasgos consiste en encontrar el subconjunto de atributos del conjunto de datos original que mejor describe los objetos del dominio; tiene como meta reducir la dimensión del conjunto de rasgos a través de la selección del subconjunto de rasgos de mejor desempeño bajo algún criterio de clasificación, durante este proceso se eliminan rasgos que podrían resultar redundantes proporcionando así una mejor representación de la información original reduciendo significativamente el costo computacional y contribuyendo a una mejor generalización del algoritmo de aprendizaje.

Una vez que las bases de datos se convierten a formato ARFF, con fines de trabajarlas en el software WEKA, quedan listas para realizar selección de rasgos.

Este proceso de selección de rasgo se les aplicó a los modelos:

- Modelo 1er año.
- Modelo 2do año.

Modelación de los Datos

Estos modelos de datos fueron confeccionados con los resultados académicos alcanzados por los estudiantes de 1er y 2do años de la carrera de Ingeniería Informática en la Universidad de Cienfuegos y constan de 16 y 18 rasgos respectivamente.

En ambos conjuntos existe un rasgo objetivo Aprueba, que toma valores en el conjunto {si,no}.

Los restantes rasgos se corresponden con los resultados alcanzados en las asignaturas del año respectivamente, y toman valores numéricos en el intervalo [0,5].

Los valores del 2 al 5 tienen el mismo significado que habitualmente se le da en el sistema de evaluación de los CES en Cuba:

2- No aprobado.

3- Regular.

4- Bien.

5- Excelente.

Y cada valor representa la nota final con que terminó esa asignatura.

Los valores 0 y 1 son utilizados para significar aquellos estudiantes que han sido Eximidos o han Convalidado alguna asignatura en un año.

Los datos fueron recolectados durante los cursos 2010-11, 2011-12, 2012-13 y 2013-14; resultando para el modelo de 1er año un total de 114 instancias y para el de segundo un total de 120.

La estructura de los modelos de datos en su formato ARFF utilizados en la selección de rasgos pueden ser consultados en el *Anexo 1*

Método

Sobre cada uno de los modelos se aplicaron los evaluadores CfsSubsetEval e InfoGain AttributeEval asociados a métodos de búsqueda y modos de evaluación como se presenta de forma resumida para cada año en las tablas siguientes.

Los resultados de la evaluación en cuanto a cantidad de atributos seleccionados por cada algoritmo utilizado se especifican en la penúltima columna de las tablas.

Resultados y discusión

Modelo 1er Año.

Tabla 1. Algoritmos de selección aplicados al modelo 1er Año.

Bases de Datos	Evaluador de atributo	Método de búsqueda	Modo Eval.	Cantidad de atributos seleccionados	Anexo
Modelo 1er año. Arff	CfsSubsetEval	BestFirst -D 1 -N 5	10 Folds CV	15	2_1
			Training set	4 (1,3,13,14)	2_2
		FCBFSearch	10 Folds CV	15	2_3
			Training set	15	2_4
	InfoGainAttributeEval	Ranker	10 Folds CV	15	2_5
			Training set	15	2_6

De la aplicación de la selección de rasgos en el Modelo de 1er Año se obtiene:

- Con diferentes evaluadores y métodos de búsqueda, que solo en uno de los seis casos se eliminan atributos. En el resto de los casos resultan seleccionados los 15 atributos. El resultado de la aplicación de los modelos puede ser consultado en el Anexo 2.(del 2.1 al 2.6).
- La combinación de CfsSubsetEval con BestFirst y modo de evaluación Training Set (ver Anexo 2.2) selecciona solo 4 atributos: Mat I, Flnf, MD y Mat II. Ciertamente, estos cuatro atributos tienen importancia en el primer año de la carrera pero no son los únicos, de los 15 atributos a discriminar quedan fuera de esta selección atributos como IP y DPOO que, por la experiencia acumulada en el colectivo de año y por valoración de los profesores de estas asignaturas, así como por entrevistas a los estudiantes tienen mucho mayor peso que el que este método les confiere. Se compara este resultado con los ofrecidos por los algoritmos

InfoGainAttributeEval con Ranker en ambos modos de evaluación y se comprueba lo aportado por los expertos y las entrevistas realizadas.(Ver Anexo2.5 y 2.6)

Se decide por el equipo investigador lo siguiente:

- Primero:* aceptar el Modelo de 1er Año con todos los atributos, teniendo en cuenta el resultado dado por 5 de los 6 métodos aplicados, y
- Segundo:* probar con un segundo modelo para el 1er Año que incluya los atributos de mayor peso (según el InfoGainAttributeEval con Ranker con ambos modos de evaluación – anexos 2.5 y 2.6) al que denominamos *Modelo_1er Año_sel_rasg* que contendrá los 10 atributos más importantes según el InfoGainAttributeEval con Ranker y modo de evaluación crossvalidation con 10 folds. Resultan seleccionados para este modelo los atributos: *Mat I, Mat II, FyS, MD, IP, DPOO, Flnf, IdEx II, Alg y GA e IGS*, listados aquí según el orden que les confiere este modelo. –ver Anexo 2.5)

Modelo 2do Año.

Tabla 2. Algoritmos de selección aplicados al modelo 2do Año.

Bases de Datos	Evaluador de atributo	Método de búsqueda	Modo Eval.	Cantidad de atributos seleccionados	Anexo
Modelo 2do año. Arff	CfsSubsetEval	BestFirst -D 1 -N 5	10 Folds CV	17	2_7
			Training set	6 (1,3,4,10,11,12)	2_8
		FCBFSearch	10 Folds CV	9	2_9
			Training set	9	2_10
	InfoGainAttributeEval	Ranker	10 Folds CV	17	2_11
			Training set	17	2_12

El resultado de la aplicación de los modelos puede ser consultado en el *Anexo 2* (del 2.7 al 2.12).

De la aplicación de la selección de rasgos en el Modelo de 2do Año se obtiene; con diferentes evaluadores y métodos de búsqueda, que:

- Cuando se utiliza CfsSubsetEval con BestFirst los resultados entre el modo de evaluación crossvalidation y training set difieren, obteniéndose en el primer caso el total de los atributos y en el segundo solo seis de ellos, siendo esta selección sometida a criterio de expertos. Se observa que en esta selección es eliminado el rasgo *Mat IV* que es considerado por todos como importante, tampoco incluye la *PProf*. (ver *Anexos* 2.7 y 2.8)
- Cuando se utiliza CfsSubsetEval con FCBFSearch los resultados convergen con ambos modos de evaluación, resultando seleccionados nueve rasgos en total, que incluye además de los seis del método anterior los rasgos *Mat IV*, *PProf* y *aprueba* que es el rasgo objetivo. (ver *Anexos* 2.9 y 2.10)
- Cuando se utiliza InfoGainAttributeEval con Ranker, se obtienen todos los rasgos ordenados por su importancia, para ambos modos de evaluación. Coincidentemente en este caso los primeros rasgos ordenados por este método coinciden con el resultado que ofrece el método anterior. (ver *Anexos* 2.11 y 2.12)
- Teniendo en cuenta lo explicado antes, se decide:
- Seleccionar como rasgos para *Modelo_2do_año_sel_rasgos* los siguientes: *mat iii*, *ed1*, *ia*, *mat iv*, *Acomp*, *ed2*, *bd*, *PProf* además del rasgo objetivo *Aprueba*. Estos son los rasgos que aporta el método CfsSubsetEval con FCBFSearch en ambos modos de evaluación.
- De los rasgos seleccionados antes se observa que los atributos eliminados por los algoritmos usados son los correspondientes a las asignaturas *Dnac*, *epcap*, *idex3*, *segNac*, *TSP*, *EF 3*, *EF iv*, *idex 4* y *elec1*. Estos rasgos corresponden con las asignaturas del ciclo de humanidades y cultura física que contemplan todos los planes de estudio de todas las carreras universitarias en el país, por lo que se decide con fines comparativos utilizar ambos modelos, el modelo original con todas las asignaturas del año (*Modelo 2do Año*), y el modelo que se obtiene como resultado de la selección de rasgos.

Una vez terminada la selección de rasgos, se tienen cuatro conjuntos de datos para explorar, tanto con redes neuronales como con árboles de decisión. Quedan entonces, los 2 conjuntos originales y los 2 que resultan de la selección de rasgos:

- Modelo 1er año_sel_rag.

- Modelo 2do año_sel_rag.

Las estructuras de los modelos obtenidos pueden ser consultados en el *Anexo 3*.

CONCLUSIONES

El proceso de selección de rasgos aplicado a los conjuntos de datos de 1er y 2do año de la carrera de Ingeniería Informática en la UCF permite concluir:

La combinación de evaluadores y métodos de búsqueda da mejores resultados que cuando se utiliza un solo método por separado.

La mejor combinación en este problema estuvo dada por el evaluador CfsSubsetEval (con FCBFSearch o BestFirst como métodos de búsqueda) unido al InfoGainAttributeEval (con Ranker).

En la selección de los rasgos influye el conocimiento del dominio y del significado de cada atributo para la investigación, lo que debe ser tenido en cuenta en conjunción con lo aportado por los algoritmos.

Los nuevos conjuntos de datos obtenidos reducen la dimensión del conjunto original, en el caso de primer año en cinco atributos (33%) y en el de segundo en nueve atributos (50%).

BIBLIOGRAFÍA

- García Morate, D. (2012). *Manual de Weka*. Recuperado de <http://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/weka.pdf>
- Weiss, S., & Indurkha, N. (1998). *Predictive data mining: A practical guide*. San Francisco: Morgan Kaufmann.
- Witten, I. H., & Eibe, F. (2005). *Data Mining: practical machine learning tools and techniques* (2da ed). San Francisco: Morgan Kaufmann Publishers.

ANEXOS

Anexo 1. Modelos originales de datos.

Anexo 1.1. Modelo 1er Año.

16 rasgos

```
@relation 'Modelo_1er año'
@attribute 'Mat I' numeric
@attribute 'FyS' numeric
@attribute 'Flnf' numeric
@attribute 'IP' numeric
@attribute 'EF I ' numeric
@attribute 'IdO I' numeric
@attribute 'Alg y GA' numeric
@attribute 'H1uba' numeric
@attribute 'IdO II' numeric
@attribute 'DPOO' numeric
@attribute 'EF II' numeric
@attribute 'IGS' numeric
@attribute 'MD' numeric
@attribute 'Mat II' numeric
@attribute 'EP1ap' numeric
@attribute 'aprueba {si,no}'
@data
5,5,4,5,5,5,4,3,4,5,5,4,4,4,5,si
3,4,4,4,5,5,3,4,4,5,5,4,3,3,3,si
4,3,4,2,5,4,3,5,3,3,5,3,3,3,4,no
3,4,4,3,5,4,4,4,3,5,5,3,3,3,3,si
3,4,4,2,5,4,3,4,3,3,5,3,3,3,4,no
3,4,4,3,5,5,3,4,3,3,5,3,3,3,4,no
3,5,5,5,5,5,4,5,3,4,5,4,3,4,4,si
3,4,4,2,5,5,3,3,3,2,5,3,3,2,4,no
2,3,4,2,5,4,3,3,3,2,5,3,3,2,4,no
2,3,4,2,5,4,2,3,3,2,5,4,2,2,3,no
3,3,5,3,5,4,3,4,3,3,5,3,3,3,3,si
5,5,4,4,5,5,4,4,3,5,5,4,4,5,4,si
3,3,4,3,5,5,3,5,3,3,5,4,3,3,4,si
3,4,5,3,5,5,4,5,4,5,5,3,3,4,4,si
```

Anexo 1.2 - Modelo 2do Año.

16 rasgos

```
@relation 'Modelo 2do año'
@attribute 'Mat III' numeric
@attribute 'Epcap' numeric
@attribute 'ED I' numeric
@attribute 'IIA' numeric
@attribute 'IdEx III' numeric
@attribute 'SegNac' numeric
@attribute 'EF 3' numeric
@attribute 'elec ' numeric
@attribute 'Mat IV' numeric
@attribute 'Arqomp' numeric
@attribute 'ED 2' numeric
@attribute 'BD' numeric
@attribute 'TSP' numeric
@attribute 'IdEx IV' numeric
@attribute 'DeNac' numeric
@attribute 'EF iv' numeric
@attribute 'Pprof' numeric
@attribute 'aprueba {no,si}'
@data
2,3,2,2,3,2,4,3,3,2,2,3,3,2,4,2,no
2,3,2,2,3,3,5,4,3,2,2,3,4,4,3,5,2,no
2,2,2,2,4,3,5,3,3,2,2,2,3,4,3,5,2,no
4,3,4,3,4,4,4,5,5,3,5,4,5,4,5,5,si
2,3,3,3,3,3,5,4,3,3,3,3,3,4,4,5,4,si
2,3,2,2,3,3,5,4,3,2,3,3,3,3,4,5,4,no
2,2,2,2,3,3,5,3,2,2,2,2,4,2,5,2,no
3,3,4,3,3,3,2,3,3,4,3,3,4,4,3,3,5,no
3,5,2,3,3,5,4,4,3,3,2,3,3,3,4,4,2,no
2,2,2,2,5,3,4,3,2,2,2,2,2,5,3,5,5,no
```

Anexo 2

Anexo 2.1

=== Run information ===

Evaluator: weka.attributeSelection.CfsSubsetEval
Search:

weka.attributeSelection.BestFirst -D 1 -N 5

Relation: Modelo_1ro año-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16

Instances: 114

Attributes: 16

Mat I

FyS

FInf

IP

EF I

IdEx I

Alg y GA

HCuba

IdEx II

DPOO

EF II

IGS

MD

Mat II

EPCap

aprueba

Evaluation mode: 10-fold cross-validation

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%) attribute

10(100 %) 1 Mat I

6(60 %) 2 FyS

10(100 %) 3 FInf

2(20 %) 4 IP

0(0 %) 5 EF I

0(0 %) 6 IdEx I

0(0 %) 7 Alg y GA

1(10 %) 8 HCuba

0(0 %) 9 IdEx II

3(30 %) 10 DPOO

0(0 %) 11 EF II

0(0 %) 12 IGS

10(100 %) 13 MD

10(100 %) 14 Mat II

0(0 %) 15 EPCap

Anexo 2.2

=== Run information ===

Evaluator: weka.attributeSelection.CfsSubsetEval
Search:

weka.attributeSelection.BestFirst -D 1 -N 5

Relation: Modelo_1ro año-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16

Instances: 114

Attributes: 16

Mat I

FyS

FInf

IP

EF I

IdEx I

Alg y GA

HCuba

IdEx II

DPOO

EF II

IGS

MD

Mat II

EPCap

aprueba

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Best first.

Start set: no attributes

Search direction: forward

Stale search after 5 node expansions

Total number of subsets evaluated: 103

Merit of best subset found: 0.414

Attribute Subset Evaluator (supervised, Class (nominal): 16 aprueba):

CFS Subset Evaluator

Including locally predictive attributes

Selected attributes: 1,3,13,14 : 4

Mat I

FInf

MD

Mat II

Anexo 2.3

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.FCBFSearch
-D false -T -1.7976931348623157E308 -N -1
Relation:  Modelo_1ro año-weka.filters.unsu-
pervised.attribute.Remove-R1-weka.filters.un-
supervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R16
Instances: 114
Attributes: 16
          Mat I
FyS
FInf
IP
          EF I
IdEx I
          Alg y GA
HCuba
IdEx II
          DPOO
          EF II
          IGS
          MD
          Mat II
EPCap
aprueba
Evaluation mode: 10-fold cross-validation
```

Anexo 2.4

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.FCBFSearch
-D false -T -1.7976931348623157E308 -N -1
Relation:  Modelo_1ro año-weka.filters.unsu-
pervised.attribute.Remove-R1-weka.filters.un-
supervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R16
Instances: 114
Attributes: 16
          Mat I
FyS
FInf
IP
          EF I
IdEx I
          Alg y GA
HCuba
IdEx II
          DPOO
          EF II
          IGS
          MD
          Mat II
EPCap
aprueba
Evaluation mode: evaluate on all training data
```

Anexo 2.5

```
=== Run information ===
Evaluator:          weka.attributeSelection.
InfoGainAttributeEval
Search:            weka.attributeSelection.Ranker -T
-1.7976931348623157E308 -N -1
Relation:         Modelo_1ro año-weka.filters.unsu-
pervised.attribute.Remove-R1-weka.filters.un-
supervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R16
Instances:        114
Attributes:        16
                  Mat I
FyS
Flnf
IP
                  EF I
IdEx I
                  Alg y GA
HCuba
IdEx II
                  DPOO
                  EF II
                  IGS
                  MD
                  Mat II
EPCap
aprueba
Evaluation mode:   10-fold cross-validation
=== Attribute selection 10 fold cross-validation
(stratified), seed: 1 ===
average merit      average rank  attribute
0.531 +- 0.028     1 +- 0       1 Mat I
0.438 +- 0.031     2 +- 0       14 Mat II
0.356 +- 0.031     3.4 +- 0.49   2 FyS
0.325 +- 0.027     4.1 +- 1.14   13 MD
0.292 +- 0.024     5 +- 0.77    4 IP
0.281 +- 0.024     5.7 +- 0.78   10 DPOO
0.239 +- 0.019     7 +- 0.63    3 Flnf
0.203 +- 0.022     8.5 +- 1.12   9 IdEx II
0.191 +- 0.019     9 +- 0.89    7 Alg y GA
0.166 +- 0.015     10.1 +- 0.83  12 IGS
0.171 +- 0.019     10.2 +- 0.98  8 HCuba
0.119 +- 0.018     12.2 +- 0.6   6 IdEx I
0.095 +- 0.013     12.9 +- 0.3   15 EPCap
0.054 +- 0.015     13.9 +- 0.3   5 EF I
0.014 +- 0.004     15 +- 0       11 EF II
```

Anexo 2.6

```
=== Run information ===
Evaluator:          weka.attributeSelection.
InfoGainAttributeEval
Search:            weka.attributeSelection.Ranker -T
-1.7976931348623157E308 -N -1
Relation:         Modelo_1ro año-weka.filters.unsu-
pervised.attribute.Remove-R1-weka.filters.un-
supervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R1-weka.filters.
unsupervised.attribute.Remove-R16
Instances:        114
Attributes:        16
                  Mat I
FyS
Flnf
IP
                  EF I
IdEx I
                  Alg y GA
HCuba
IdEx II
                  DPOO
                  EF II
                  IGS
                  MD
                  Mat II
EPCap
aprueba
Evaluation mode:   evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
                  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal):
16 aprueba):
                  Information Gain Ranking Filter

Ranked attributes:
0.5274  1 Mat I
0.4347  14 Mat II
0.3479  2 FyS
0.3213  13 MD
0.2845  4 IP
0.2743  10 DPOO
0.2337  3 Flnf
0.198   9 IdEx II
0.1852  7 Alg y GA
0.1657  8 HCuba
0.1603  12 IGS
0.112   6 IdEx I
0.0861  15 EPCap
0.0444  5 EF I
0.0102  11 EF II
Selected attributes: 1,14,2,13,4,10,3,9,7,8,12,6,15
,5,11 : 15
```

Anexo 2.7

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.BestFirst -D
1 -N 5
Relation:  Modelo 2do año-weka.filters.unsuper-
vised.attribute.Remove-R1-4,22
Instances: 120
Attributes: 18
            mat iii
epcap      ed1
            iia
idex 3
segNac     EF 3
elec 1     mat iv
Acomp      ED 2
            BD
            TSP
idex 4
Dnac       EF iv
Pprof
aprueba
Evaluation mode: 10-fold cross-validation
=== Attribute selection 10 fold cross-validation
(stratified), seed: 1 ===
number of folds (%) attribute
10(100 %) 1 mat iii
1( 10 %) 2 epcap
10(100 %) 3 ed1
10(100 %) 4 iia
0( 0 %) 5 idex 3
0( 0 %) 6 segNac
4( 40 %) 7 EF 3
0( 0 %) 8 elec 1
4( 40 %) 9 mat iv
5( 50 %) 10 Acomp
10(100 %) 11 ED 2
10(100 %) 12 BD
0( 0 %) 13 TSP
0( 0 %) 14 idex 4
0( 0 %) 15 Dnac
0( 0 %) 16 EF iv
3( 30 %) 17 Pprof
```

Anexo 2.8

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.BestFirst -D
1 -N 5
Relation:  Modelo 2do año-weka.filters.unsuper-
vised.attribute.Remove-R1-4,22
Instances: 120
Attributes: 18
            mat iii
epcap      ed1
            iia
idex 3
segNac     EF 3
elec 1     mat iv
Acomp      ED 2
            BD
            TSP
idex 4
Dnac       EF iv
Pprof
aprueba
Evaluation mode: evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 136
Merit of best subset found: 0.412

Attribute Subset Evaluator (supervised, Class (no-
minal): 18 aprueba):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 1,3,4,10,11,12 : 6
            mat iii
            ed1
            iia
Acomp      ED 2
            BD
```

Anexo 2.9

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.FCBFSearch
-D true -T -1.7976931348623157E308 -N -1
Relation:  Modelo 2do año-2clases-weka.filters.
unsupervised.attribute.Remove-R2,5-8,13-16
Instances: 120
Attributes: 9
          mat iii
          ed1
iia
          mat iv
          A1omp
          ED 2
          BD
Pprof
          aprueba2
Evaluation mode: 10-fold cross-validation
```

Anexo 2.10

```
=== Run information ===
Evaluator: weka.attributeSelection.CfsSubsetEval
Search:    weka.attributeSelection.FCBFSearch
-D true -T -1.7976931348623157E308 -N -1
Relation:  Modelo 2do año-2clases-weka.filters.
unsupervised.attribute.Remove-R2,5-8,13-16
Instances: 120
Attributes: 9
          mat iii
          ed1
iia
          mat iv
          A1omp
          ED 2
          BD
Pprof
          aprueba2
Evaluation mode: evaluate on all training data
```

Anexo 2.11

```

=== Run information ===
Evaluator:          weka.attributeSelection.
InfoGainAttributeEval
Search:            weka.attributeSelection.Ranker -T
-1.7976931348623157E308 -N -1
Relation:         Modelo 2do año-weka.filters.unsuper-
vised.attribute.Remove-R1-4,22
Instances:        120
Attributes:        18
                  mat iii
epcap
                  ed1
iaa
idex 3
segNac
                  EF 3
elec 1
                  mat iv
Acomp
                  ED 2
                  BD
                  TSP
idex 4
Dnac
                  EF iv
Pprof
aprueba
Evaluation mode:   10-fold cross-validation
=== Attribute selection 10 fold cross-validation
(stratified), seed: 1 ===
average merit      average rank attribute
0.475 +- 0.028     1.1 +- 0.3   3 ed1
0.403 +- 0.028     2.6 +- 1.11  1 mat iii
0.4 +- 0.032       2.9 +- 0.7   12 BD
0.383 +- 0.027     3.6 +- 0.66  11 ED 2
0.329 +- 0.019     5.7 +- 1     9 mat iv
0.315 +- 0.038     6.4 +- 1.69  4 iia
0.314 +- 0.025     6.9 +- 1.14  10 Acomp
0.299 +- 0.015     7.5 +- 1.12  15 Dnac
0.291 +- 0.027     8.3 +- 0.9   2 epcap
0.23 +- 0.029      10.8 +- 0.98  17 Pprof
0.226 +- 0.028     10.9 +- 0.94  5 idex 3
0.198 +- 0.021     12.8 +- 1.33  6 segNac
0.196 +- 0.028     13.2 +- 1.17  13 TSP
0.178 +- 0.03      13.7 +- 2.28  7 EF 3
0.175 +- 0.026     14.4 +- 0.92  16 EF iv
0.156 +- 0.014     15.6 +- 0.8   14 idex 4
0.133 +- 0.014     16.6 +- 0.92  8 elec 1

```

Anexo 2.12

```

=== Run information ===
Evaluator:          weka.attributeSelection.
InfoGainAttributeEval
Search:            weka.attributeSelection.Ranker -T
-1.7976931348623157E308 -N -1
Relation:         Modelo 2do año-weka.filters.unsuper-
vised.attribute.Remove-R1-4,22
Instances:        120
Attributes:        18
                  mat iii
epcap
                  ed1
iaa
idex 3
segNac
                  EF 3
elec 1
                  mat iv
Acomp
                  ED 2
                  BD
                  TSP
idex 4
Dnac
                  EF iv
Pprof
aprueba
Evaluation mode:   evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:Attribute ranking.

Attribute Evaluator (supervised, Class (nominal):
18 aprueba):
Information Gain Ranking Filter
Ranked attributes:
0.468  3 ed1
0.396  12 BD
0.394  1 mat iii
0.38  11 ED 2
0.323  9 mat iv
0.312  10 Acomp
0.303  4 iia
0.288  15 Dnac
0.279  2 epcap
0.222  17 Pprof
0.215  5 idex 3
0.186  13 TSP
0.185  6 segNac
0.17  7 EF 3
0.167  16 EF iv
0.148  14 idex 4
0.124  8 elec 1
Selected attributes: 3,12,1,11,9,10,4,15,2,17,5,13
,6,7,16,14,8 : 17

```

Anexo 3

Anexo 3.1. Modelo 1er año con selección de rasgos.

11 rasgos

```
@relation 'Modelo_1er año selrasg'
@attribute 'Mat I' numeric
@attribute FyS numeric
@attribute Flnf numeric
@attribute IP numeric
@attribute 'Alg y GA' numeric
@attribute 'Id0 II' numeric
@attribute DPOO numeric
@attribute IGS numeric
@attribute MD numeric
@attribute 'Mat II' numeric
@attribute aprueba {si,no}
@data
5,5,4,5,4,4,5,4,4,4,si
3,4,4,4,3,4,5,4,3,3,si
4,3,4,2,3,3,3,3,3,3,no
3,4,4,3,4,3,5,3,3,3,si
3,4,4,2,3,3,3,3,3,3,no
3,4,4,3,3,3,3,3,3,3,no
3,5,5,5,4,3,4,4,3,4,si
3,4,4,2,3,3,2,3,3,2,no
2,3,4,2,3,3,2,3,3,2,no
3,5,5,3,3,3,4,4,3,5,si
3,5,5,3,4,4,3,3,3,4,si
3,4,4,3,3,3,3,4,3,4,no
4,3,4,2,3,3,3,3,3,3,no
3,4,4,3,4,3,5,3,3,3,si
3,4,4,2,3,3,3,3,3,3,no
```

Anexo 3.2. Modelo 2do año con selección de rasgos.

9 rasgos

```
@relation 'Modelo 2do año selrasg'
@attribute 'mat iii' numeric
@attribute ed1 numeric
@attribute iia numeric
@attribute 'mat iv' numeric
@attribute A1omp numeric
@attribute 'ED 2' numeric
@attribute BD numeric
@attribute Pprofnumeric
@attribute aprueba {no,si}
@data
2,2,2,3,3,2,2,2,no
2,2,2,3,2,2,3,2,no
2,2,2,3,2,2,2,2,no
4,4,3,5,3,5,4,5,si
2,3,3,3,3,3,3,4,si
2,2,2,3,2,3,3,4,no
4,4,3,5,3,5,4,5,si
2,2,2,3,3,2,2,2,no
5,3,4,5,4,4,5,5,si
3,3,3,4,4,4,4,5,si
3,3,3,4,3,3,4,4,si
5,5,5,5,5,5,5,5,si
3,3,3,4,3,3,4,5,si
3,3,2,3,3,2,4,4,no
5,3,3,4,3,4,4,4,si
4,3,3,5,4,3,5,5,si
4,3,4,4,3,3,5,4,si
```