

Las máquinas aprenden, los alumnos también

Dr. Francisco Pérez García

Jefe del Departamento de Tecnología del Instituto Pompeu Fabra de Martorell

Profesor Asociado de Farmacología de la Universidad de Barcelona

Email: perez@ub.edu. Web: <http://www.technologies.net>

Resumen

El objetivo del presente trabajo es dar a conocer una experiencia educativa aplicada en enseñanza secundaria para demostrar que los alumnos pueden aprender a enseñar aprendizaje automático. Las herramientas informáticas empleadas para el aprendizaje automático han sido Weka (Universidad de Waikato, Nueva Zelanda), RStudio y ml5.js (Fundación Processing). Mediante el uso de las herramientas anteriores se han analizado conjuntos de datos clásicos como Iris o Flavia. Iris contiene 150 datos de amplitud y longitud de pétalos y sépalos para la identificación de 3 especies vegetales del género Iris. Flavia contiene 1892 imágenes de 32 hojas de otras tantas especies vegetales. Una vez entendido estos y otros ejemplos los alumnos lo han aplicado a la identificación automática de imágenes pertenecientes a diferentes clases.

Palabras clave: Machine learning, pensamiento computacional, creatividad, tecnología

1. Introducción

En estudios anteriores hemos destacado la importancia de la investigación en la enseñanza de la tecnología (Pérez-García 2001). Para ello es necesario una actualización del profesorado en temas informáticos (Pérez-García, 2010) y especialmente en temas relacionados con Arduino y lenguajes de programación. La programación en diferentes lenguajes informáticos permite la creatividad tecnológica de nuestros alumnos. Se presentará la situación en España y otros países por lo que se refiere al empleo de robótica educativa, creaciones a partir de la plataforma Arduino y las posibilidades que ofrece Processing para crear objetos interactivos de cierta complejidad (Pérez-García, 2013, 2014). Además el lenguaje Javascript presenta grandes posibilidades educativas innovadoras desde la creación de juegos y la unión simultánea a internet de las cosas hasta realidad aumentada, virtual, y cualquier cosa que podamos imaginar y de la que exista una biblioteca Javascript y si no existiese es factible crearla (Pérez-García, 2016).

En el presente trabajo se expondrán herramientas educativas como por ejemplo Weka, R, RWeka, ml.js, ml5.js que permiten a los alumnos que ellos pueden enseñar a las máquinas.

El aprendizaje automático se sitúa dentro de una posible secuencia de actividades a realizar por los alumnos:

- a) Programación de bloques con Scratch y/o MIT App Inventor
- b) Programación básica con HTML5, CSS y Javascript (w3schools.com)
- c) Programación interactiva básica con teclado, imágenes, videos, sonidos con p5.js (p5js.org)

- d) Programación interactiva conectando Arduino con una web del alumno con librería Javascript johnny-five.js y datos propios en JSON
- e) Programación con ml5.js, R y weka sobre aprendizaje automático

2. El aprendizaje de las máquinas

Arthur Samuel, un pionero en inteligencia artificial en IBM y Stanford, definió el aprendizaje automático de la siguiente manera:

El campo de estudio que le da a los ordenadores la capacidad de aprender sin ser programado explícitamente

Las herramientas empleadas por los alumnos para conocer, aplicar y crear conocimiento a partir de datos son Weka, ml5.js y R (Figura 1).



Figura 1. Herramientas informáticas para el aprendizaje automático

El aprendizaje automático es un campo tecnológico clave para el futuro de la humanidad en la que han trabajado y trabajan investigadores en todos y cada uno de los campos del conocimiento humano y en la que se ha producido avances informáticos durante las últimas décadas.

Fundamentalmente, el aprendizaje automático utiliza algoritmos para extraer información de datos sin procesar y representarla en algún tipo de modelo. Utilizamos este modelo para inferir cosas sobre otros datos que aún no hemos modelado.

En los últimos años el crecimiento del poder computacional exponencial ha hecho que la industria viera una explosión de técnicas computacionales que antes no era posible.

El aprendizaje automático consiste en el estudio de algoritmos y modelos estadísticos que utilizan los ordenadores para mejorar su rendimiento en una tarea específica (por lo tanto

el uso de la palabra “aprendizaje”). Estos algoritmos utilizan lo que se conoce como “datos de entrenamiento”, un modelo matemático de datos de muestra.

La mayoría de los programas de aprendizaje automático se puede dividir en dos categorías:

- 1) aprendizaje supervisado es aquel que utiliza conjuntos de datos que han sido previamente etiquetados como ejemplos para ser entrenado en él.

- 2) aprendizaje no supervisado es aquel que utiliza un conjunto de datos que no ha sido etiquetado o categorizado.

Los alumnos aprenden con ejemplos contenidos en los programas Weka, R y ml5.js y posteriormente los aplican a los problemas que eligen. Se enseña en lengua inglesa con metodología AICLE.

3. Weka

Weka (Waikato Environment for Knowledge Analysis, en español «entorno para análisis del conocimiento de la Universidad de Waikato») es una plataforma de software para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato.

Weka se puede descargar desde su página web oficial (Weka, 2019). Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, agrupación, extracción de reglas de asociación y visualización.

Weka es un ave que no vuela y que se considera que tiene una naturaleza inquisitiva y que es endémico de las islas de Nueva Zelanda. Weka es un software de código abierto emitido bajo la Licencia Pública General de GNU.

La plataforma futurelearn reúne varios cursos gratuitos en línea que enseñan el aprendizaje automático y la minería de datos utilizando Weka a nivel introductorio (Futurelearn, 2019a) y a nivel avanzado (Futurelearn, 2019b). Además, los videos para los cursos están disponibles en Youtube. Sí, es posible aplicar Weka para procesar *big data* y realizar un aprendizaje profundo en enseñanza secundaria.

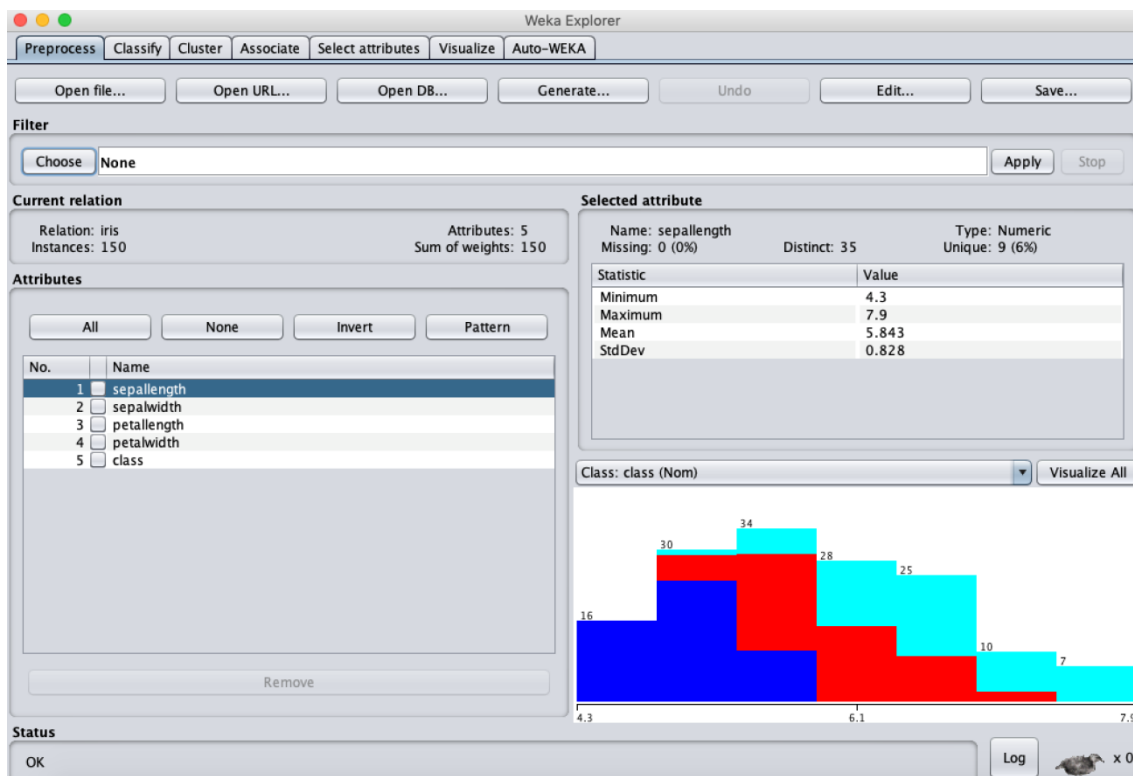


Figura 2. Ejemplo de datos de longitud y ancho de pétalos y sépalos de 3 especies de Iris para aprendizaje automático con Weka.

La potencia de Weka en el aprendizaje automático es muy alta y no requiere conocimientos de programación. Permite obtener y analizar datos numéricos con numerosos algoritmos de aprendizaje automático que proceden de algoritmos como filtros que dan lugar a atributos que son datos numéricos generados a partir de imágenes, pero también se pueden obtener a partir de cualquier tipo de dato (sonidos, videos, textos, cualquier tipo de big data). Weka selecciona el mejor algoritmo para aprender y hace cálculos de su eficacia a partir de la matriz de confusión (Figura 3).

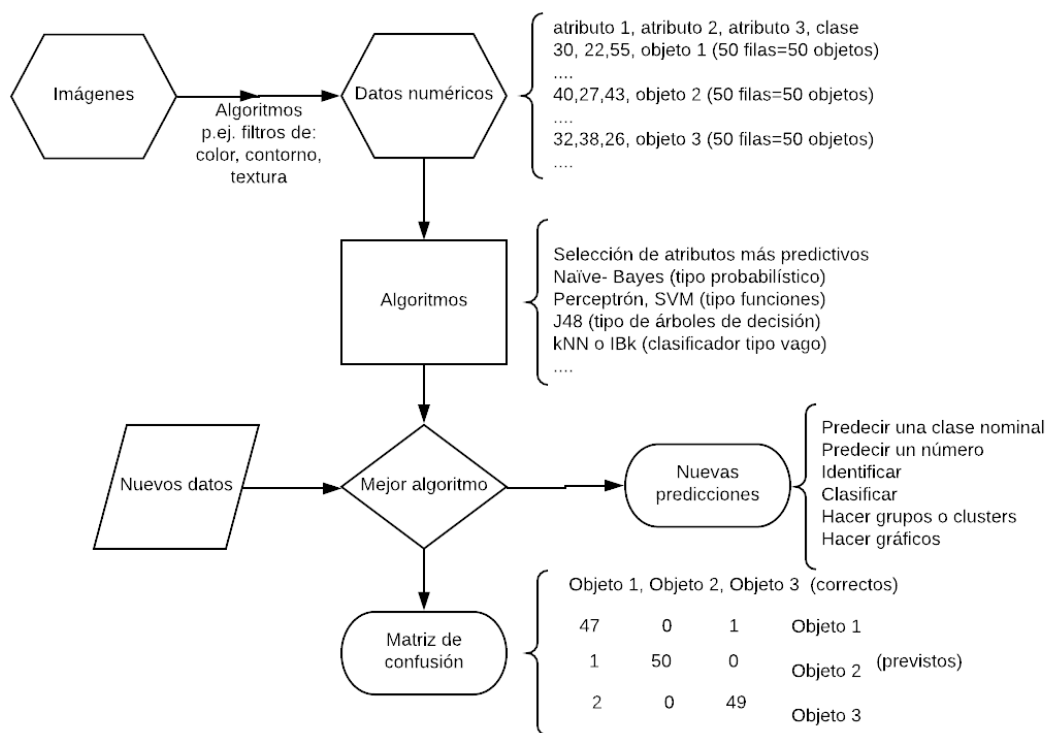


Figura 3. Flujo de trabajo a partir de imágenes o datos numéricos en aprendizaje por máquina empleando Weka.

Weka no solo incluye numerosos algoritmos de aprendizaje sino que incluye muchos ejemplos de bases de datos en el propio programa, como por ejemplo la base de datos de flores de Iris (https://es.wikipedia.org/wiki/Iris_flor_conjunto_de_datos)

4. Lenguaje R y RStudio

R es un lenguaje de programación sencillo ampliamente empleado en análisis de datos y en aprendizaje automático. Entre las bibliotecas informáticas complementarias de R hay algunas muy empleadas como por ejemplo “caret” (CARET, 2018). El paquete caret de R es la abreviatura de *Classification And REgression Training* y consiste en un

conjunto de funciones que intentan agilizar el proceso para crear modelos predictivos. El paquete `caret` contiene herramientas para:

- división de datos
- pre-procesamiento
- selección de características
- ajuste del modelo mediante remuestreo
- estimación de importancia variable
- otras funcionalidades

Otras bibliotecas o paquetes de R empleadas en aprendizaje automático son “`rpart`” (RPART, 2018) o “`party`”. Ambas librerías tiene implementados diferente algoritmos con árboles de decisión, entre otros el algoritmo `cree`, un árbol de regresión (PARTY, 2018).

A modo de ejemplo de clustering se expone el caso clásico de clasificación de las 3 especies de Iris con R (Figura 4 y 5)

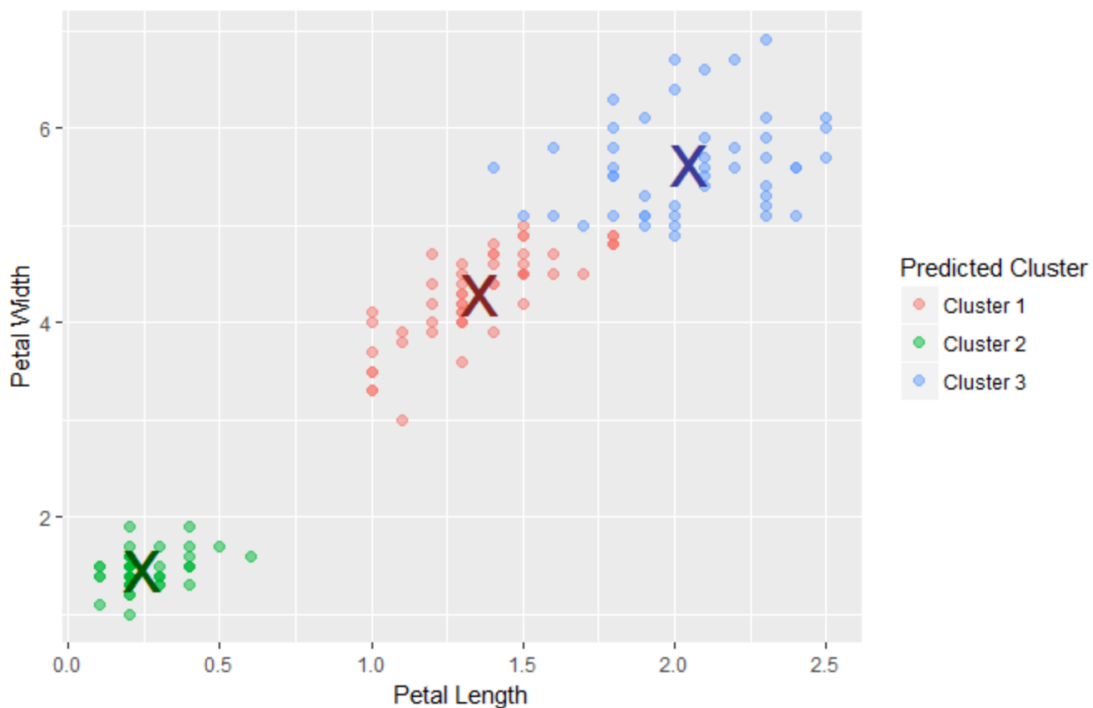


Figura 4. Ejemplo de gráfico con R de agrupamiento o clustering con datos de 3 especies de flor de género Iris.

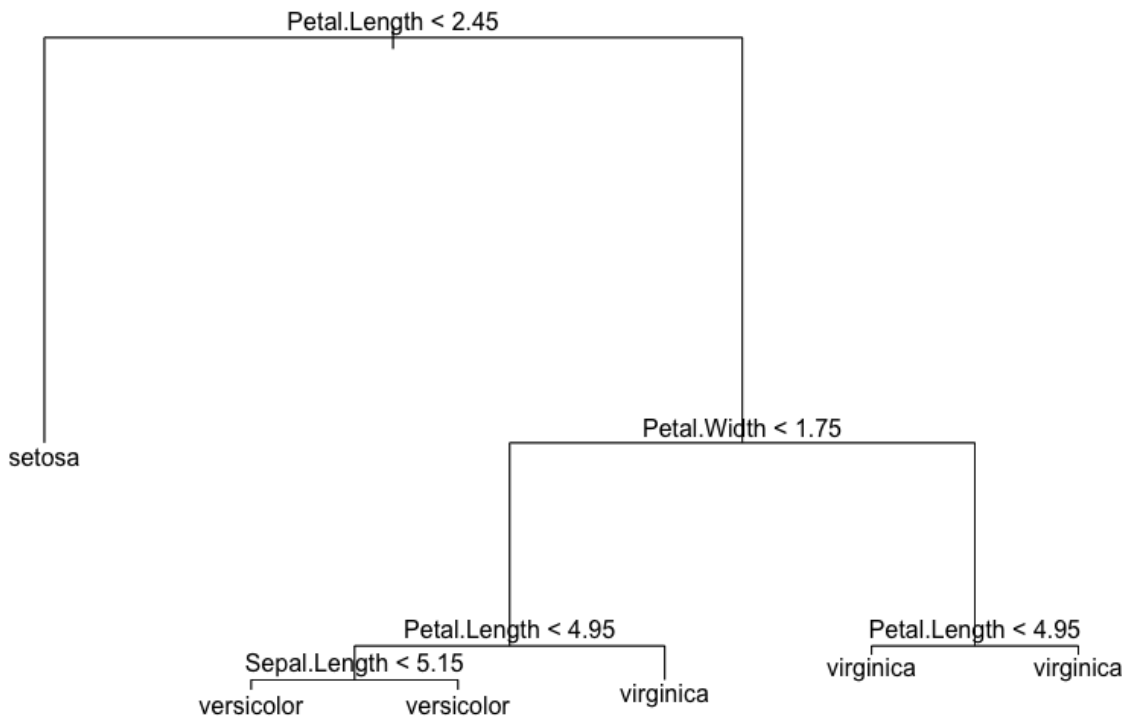


Figura 5. Ejemplo de árbol de decisión obtenido a partir de datos del género Iris.

5. ml5.js

ml5.js tiene como objetivo hacer que el aprendizaje automático sea accesible para cualquier estudiante. La biblioteca proporciona acceso a modelos y algoritmos de aprendizaje automático en el navegador, construidos sobre TensorFlow.js sin otras dependencias externas y está disponible en www.ml5js.org. La biblioteca es compatible con ejemplos de código, tutoriales y conjuntos de datos de muestra con un énfasis en informática ética y de código abierto.

A partir de los ejemplos disponibles online se ha partido de un ejemplo que reconoce piedra, papel o tijera que es capaz de entrenarse con imágenes del usuario y guardar los datos de entrenamiento en formato json (Figura 6) para reutilizarse con 3 imágenes cualesquiera tal como explica el video de uno de los líderes de ml5 (Shiffman, 2019).

Cada alumno ha trabajado con imágenes diferentes, por ejemplo imágenes laterales de diferentes modelos de coches de una marca determinada o de diferentes fotografías de especies vegetales o animales. Se presentan resultados de aprendizaje automáticos realizados por el alumno Guillermo Raya con la base de datos Flavia de 32 especies vegetales.



FeatureExtractor(mobileNet model) Loaded

Load Dataset

If you load this sample classifier dataset. Try to make rock, paper, or scissor gestures to see if the classifier can class them. If this sample dataset doesn't work well for you, you could train your own classifier, and use the 'Save Dataset' button below to create your own myKNNDataset.json file, and replace the myKNNDataset.json in this folder.'



Add an Example to Class Rock

Reset Class Rock

35 Rock examples | Confidence in Rock is: 100 %



Add an Example to Class Paper

Reset Class Paper

35 Paper examples | Confidence in Paper is: 0 %



Add an Example to Class Scissor

Reset Class Scissor

37 Scissor examples | Confidence in Scissor is: 0 %

Start predicting!

Clear all classes

KNN Classifier with mobileNet model labeled this as Class: Rock with a confidence of 100 %

Save Dataset

Figura 6. Ejemplo de autoaprendizaje con ml5.js reutilizable a partir de los videotutoriales de Shifman, 2019 disponible en www.tecnologies.net/KNN para su reutilización.



Figura 7. Treinta y dos hojas procedentes de la base de datos Flavia (1892 fotos) analizadas con R, Weka y ml5.js

En las figuras 8 y 9 se pueden observar los resultados de los algoritmos de aprendizaje J48 obtenidos a partir de todas imágenes de hojas

```

1. J48 pruned tree
2. -----
3.
4. Gabor Features0 <= 14.740877
5. |   FCTH86 <= 0
6. |   |   FCTH38 <= 0
7. |   |   |   FCTH14 <= 1
8. |   |   |   |   PHOG363 <= 3
9. |   |   |   |   |   PHOG384 <= 0
10. |   |   |   |   |   |   FCTH12 <= 1: Prunus-persica (3.0)
11. |   |   |   |   |   |   |   FCTH12 > 1: Indigofera-tinctoria-L (3.0)
12. |   |   |   |   |   |   |   |   PHOG384 > 0: Liriodendron-chinense (3.0/1.0)
13. |   |   |   |   |   |   |   |   |   PHOG363 > 3
14. |   |   |   |   |   |   |   |   |   |   PHOG159 <= 0: Lagerstroemia-indica (50.0/1.0)
15. |   |   |   |   |   |   |   |   |   |   |   PHOG159 > 0
16. |   |   |   |   |   |   |   |   |   |   |   |   PHOG376 <= 2: Ligustrum-lucidum (29.0)
17. |   |   |   |   |   |   |   |   |   |   |   |   |   PHOG376 > 2
18. |   |   |   |   |   |   |   |   |   |   |   |   |   |   FCTH14 <= 0: Ginkgo-biloba (3.0/1.0)
19. |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   FCTH14 > 0: Kalopanax-septemlobus (2.0)
20. |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   FCTH14 > 1
    
```

Figura 8. Parte del árbol de decisión J48 creado para la identificación de hojas de 32 especies vegetales


```

===348 Confusion Matrix ===
1.  a b c d e f g h i j k l m n o p q r s t u v w x y z aa ab ac ad ae af <-- classified as
2.  47 0 0 0 0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 2 3 0 0 1 0 3 0 0 0 0 0 0 | a = Phyllostachys-edulis
3.  0 42 0 0 0 0 0 3 2 1 0 1 0 0 4 0 1 0 1 0 0 0 2 0 3 1 0 1 0 0 0 1 0 | b = Aesculus-chinensis
4.  0 0 56 0 0 0 0 1 0 0 0 2 0 1 2 0 2 0 2 2 0 0 0 0 0 0 0 0 1 0 1 1 1 0 | c = Cercis-chinensis
5.  1 1 0 64 0 0 0 0 0 1 2 0 2 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 | d = Indigofera-tinctoria-L
6.  0 0 0 0 0 47 0 4 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 | e = Acer-palmatum
7.  0 0 0 0 0 0 49 0 0 3 0 1 2 0 0 0 0 0 0 0 0 0 4 0 0 0 0 1 1 1 0 0 0 0 0 | f = Phoebe-nanmu
8.  0 2 1 0 0 0 0 42 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 0 3 0 0 0 0 0 | g = Kalopanax-septemlobus
9.  0 2 0 0 0 0 0 0 52 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 2 0 0 0 0 0 0 0 | h = Koelreuteria-paniculata
10. 1 4 0 1 0 3 0 0 0 33 1 1 3 1 0 0 0 0 0 0 0 4 1 0 0 0 0 2 0 0 0 0 0 0 0 | i = Cinnamomum-japonicum
11. 2 0 0 2 0 0 0 0 0 2 59 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | j = Berberis-anhweiensis
12. 0 3 0 0 0 0 0 1 0 1 0 34 0 0 3 0 0 0 2 2 0 0 1 0 0 0 0 0 1 0 1 1 0 1 0 | k = Ilex-macrocarpa
13. 0 0 0 2 0 1 0 0 0 1 0 2 44 0 1 0 1 0 0 0 0 0 0 0 1 0 0 1 5 2 1 0 1 1 | l = Pittosporum-tobira
14. 0 3 0 0 0 0 0 1 0 0 0 5 0 28 0 0 2 0 0 1 0 0 4 0 0 2 0 0 2 0 0 1 3 0 0 | m = Chimonanthus-praecox-L
15. 0 6 2 0 0 0 0 0 1 0 0 0 0 0 46 1 2 0 0 0 0 0 1 0 0 0 0 1 2 0 3 0 0 0 | n = Cinnamomum-camphora
16. 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 50 0 0 0 0 0 0 0 0 0 0 0 0 1 5 2 0 0 0 | o = Viburnum-awabuki
17. 0 3 2 1 0 1 0 0 0 0 1 2 1 0 0 38 0 0 0 0 0 2 0 0 1 0 0 0 0 0 0 0 0 4 0 | p = Osmanthus-fragrans
18. 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 76 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | q = Cedrus-deodara
19. 0 2 0 0 1 0 0 0 0 1 0 2 0 3 0 0 0 0 0 50 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 | r = Ginkgo-biloba
20. 0 0 0 1 0 0 0 0 0 0 0 3 0 2 0 0 0 0 0 53 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 | s = Lagerstroemia-indica
21. 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 54 5 0 0 0 1 3 0 0 0 0 0 0 0 0 0 | t = Nerium-oleander
22. 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 2 53 0 0 0 0 0 0 0 0 0 0 0 0 | u = Podocarpus-macrophyllus
23. 1 4 1 0 0 0 0 0 0 2 0 1 2 5 1 0 1 0 0 0 0 0 32 0 0 2 0 0 2 0 0 1 0 0 | v = Prunus-serrulata
24. 0 0 2 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 46 0 0 0 0 2 0 0 1 0 | w = Ligustrum-lucidum
25. 2 5 0 3 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 46 4 2 0 0 0 0 0 0 0 0 | x = Toona-sinensis
26. 0 2 0 0 0 0 2 0 4 1 0 0 1 1 0 0 1 0 0 0 0 1 1 6 31 0 0 0 0 0 0 0 3 0 | y = Prunus-persica
27. 3 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 2 0 1 1 0 38 1 0 1 0 0 0 0 | z = Manglietia-fordiana
28. 0 0 2 0 6 1 5 1 2 0 0 0 0 0 1 0 1 0 2 0 0 1 0 0 1 1 0 28 0 0 0 1 0 0 | aa = Acer-buergerianum
29. 0 0 0 0 0 0 0 0 0 1 0 1 4 2 5 0 1 0 0 0 0 0 0 0 0 0 0 0 40 0 1 0 0 0 | ab = Mahonia-bealei
30. 0 0 0 0 0 0 0 0 0 0 0 0 4 0 1 6 2 0 0 0 0 0 0 0 0 0 0 0 1 0 42 1 0 0 | ac = Magnolia-grandiflora
31. 0 1 0 0 0 0 0 0 0 0 0 0 2 2 1 0 0 0 0 0 0 0 0 2 0 0 0 0 2 0 53 1 0 0 | ad = Populus-x-canadensis
32. 0 3 1 0 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 2 0 1 0 1 1 0 3 35 0 0 | ae = Liriodendron-chinense
33. 0 1 1 0 0 0 0 0 0 0 0 0 2 0 2 0 0 0 0 0 3 0 0 1 1 0 2 0 0 0 43 0 0 | af = Citrus-reticulata
    
```

Figura 9. Matriz de confusión de las 32 especies vegetales de la base de datos Flavia

Actualmente los alumnos trabajan con la base de datos PlantCLEF2017 que contienen en 20 Gigabytes de datos 260000 imágenes que corresponden a 10000 especies vegetales para intentar encontrar el mejor algoritmo para identificarlas ya que los programas empleados son profesionales y permiten el trabajo con “big data” de forma relativamente sencilla.

Se ha implementado en la página web www.tecnologies.net/leavesClassifier un algoritmo de aprendizaje automático basado en kNN, que es un algoritmo vago no muy difícil de implementar con la biblioteca ml5.js (Figura 9). Recuerde dar permiso a la cámara, cargar la base de datos json con el aprendizaje hecho y clicar a “Start predicting” mostrando una imagen de una de las hojas encontrada físicamente o en internet.

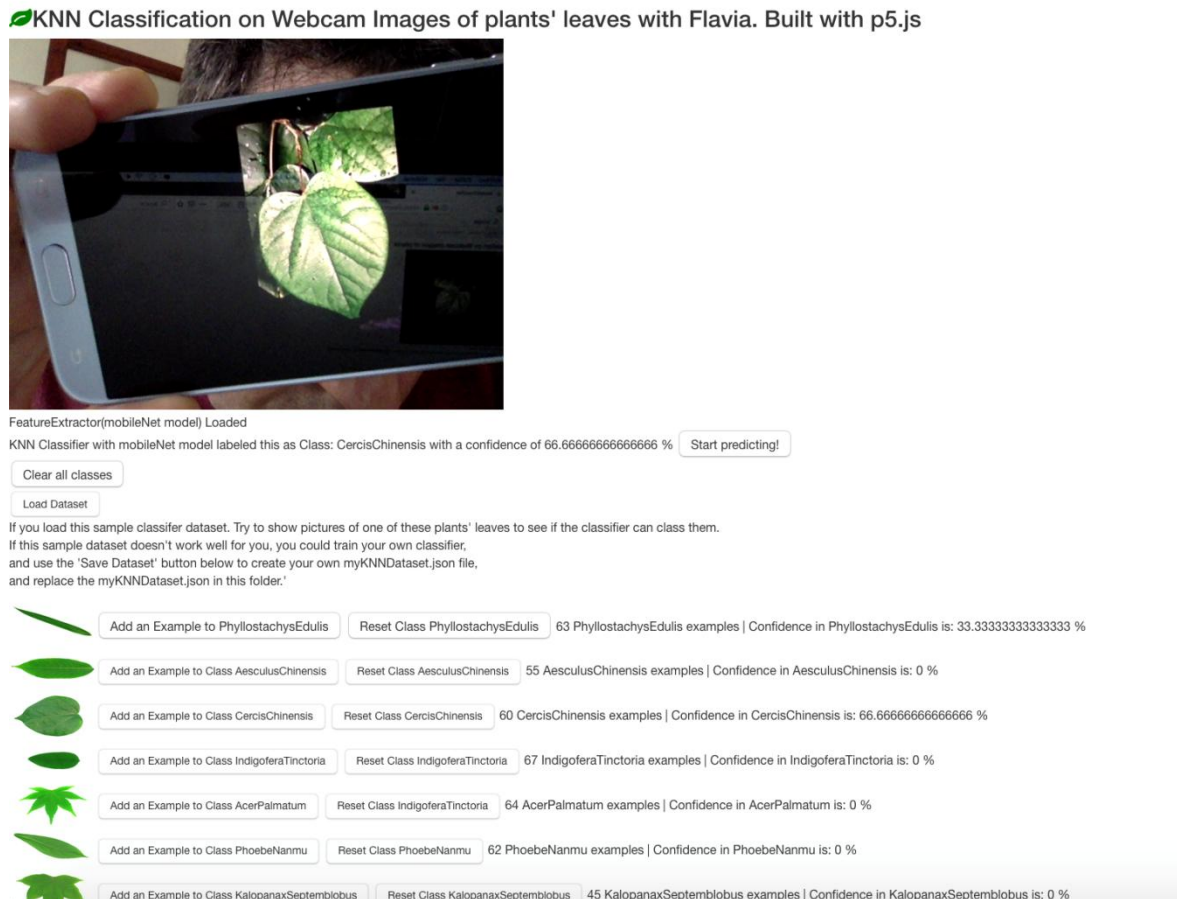


Figura 10. Identificación de 32 especies vegetales en tiempo real basado en algoritmo kNN realizado a partir de la base de datos Flavia

Agradecimientos

Al programa Erasmus+ por la concesión del proyecto 2017-1-DE03-KA219-035699_2 y el proyecto 2017-1-TR01-KA219-045575_3-4.

Al Departament d' Ensenyament de la Generalitat de Catalunya por la concesión del proyecto Technological Creativity de enseñanza AICLE (<http://tecnologies.net>).

Referencias

CARET (2019). <http://topepo.github.io/caret/index.html>

FUTURELEARN (2019a). Data mining with Weka. The University of Waikato. <https://www.futurelearn.com/courses/data-mining-with-weka>

FUTURELEARN (2019b). Advanced data mining with Weka. The University of Waikato. <https://www.futurelearn.com/courses/advanced-data-mining-with-weka>

PARTY (2018). Party package for R. <https://cran.r-project.org/web/packages/party/party.pdf>

PÉREZ GARCÍA, Francisco (2001). Exposició del treballs de recerca en tecnologia mitjançant pòsters científics. En: Tecnologia a l'ensenyament secundari. Societat Catalana de Tecnologia. Institut d'Estudis Catalans.

PÉREZ GARCÍA, Francisco (2010). Alfabetización digital del profesorado: herramientas educativas Interactivas. En Revista Didáctica, Innovación y Multimedia, núm. 16 <http://dim.pangea.org/revistaDIM16/docs/francescperez.pdf>

PÉREZ GARCÍA, Francisco (2013). Arduino en el bachillerato tecnológico, En Revista Didáctica, Innovación y Multimedia, núm. 25 <http://dim.pangea.org/revistaDIM25/revista25OCarduino.htm>

PÉREZ GARCÍA, Francisco (2014). Creatividad tecnológica mediante programación En Revista Didáctica, Innovación y Multimedia, núm. 30 <http://dim.pangea.org/revistaDIM30/docs/OC30programar.pdf>

PÉREZ GARCÍA, Francisco (2016). Javascript com a eina educativa. JORNADA SOBRE INNOVACIÓ A L'ENSENYAMENT DE LA TECNOLOGIA, pag. 98 a 108. Ed. CESIRE. Departament d'Ensenyament, IEC-SCT. <https://drive.google.com/file/d/0B-5y7UkJpGANWWWhRUDJBdzNkTmc/view>

RPART (2018). RPart package for R. <https://cran.r-project.org/web/packages/rpart/rpart.pdf>

SHIFFMAN, D. (2019). <http://shiffman.net>

WEKA, página oficial <https://www.cs.waikato.ac.nz/ml/weka/>