

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

CENTRO UNIVERSITARIO UAEM NEZAHUALCÓYOTL

**MANUAL PARA PRÁCTICAS DE LA UNIDAD DE APRENDIZAJE
TALLER CON WEKA**



**ELABORARÓN:
DR. RICARDO RICO MOLINA
DRA. CARMEN LILIANA RODRÍGUEZ PÁEZ
M. en A. VÍCTOR MANUEL DURAN LOPEZ**

AGOSTO 2019

MANUAL PARA PRÁCTICAS DE LA UNIDAD DE APRENDIZAJE TALLER CON WEKA

IDENTIFICACIÓN DE LA UNIDAD DE APRENDIZAJE

ESPACIO ACADÉMICO: Centro Universitario Nezahualcóyotl							
PROGRAMA EDUCATIVO: LICENCIATURA EN INGENIERÍA EN SISTEMAS INTELIGENTES					Área de docencia: Sustantiva		
Aprobación por los H.H. Consejos Académico y de Gobierno				Fecha: Agosto 2019		Programa elaborado por: Dr. Ricardo Rico Molina	
Nombre de la Unidad de Aprendizaje: TALLER CON WEKA					Fecha de elaboración: AGOSTO 2019		
Clave L40661	Horas de teoría 1	Horas de práctica 2	Total, de horas 3	Créditos 4	Tipo de Unidad de Aprendizaje Curso Taller	Carácter de la Unidad de Aprendizaje Obligatoria	Núcleo de formación Sustantivo
Prerrequisitos		Unidad de Aprendizaje Antecedente Ninguna			Unidad de Aprendizaje Consecuente Ninguna		

EL PRESENTE MANUAL DE PRÁCTICAS HA SIDO AVALADO EN EL MES DE SEPTIEMBRE DE 2019 POR:

  	
M. EN C. JOSÉ A. CASTILLO JIMENEZ SECRETARIO DEL H. CONSEJO	M. EN C. JOSÉ A. CASTILLO JIMENEZ SECRETARIO DEL H. CONSEJO
H. CONSEJO DE GOBIERNO CENTRO UNIVERSITARIO NEZAHUALCÓYOTL	H. CONSEJO ACADÉMICO CENTRO UNIVERSITARIO NEZAHUALCÓYOTL

ÍNDICE

Directorio UAEM	5
Directorio del Centro UAEM Nezahualcóyotl	6
Ubicación de la asignatura Taller de Weka dentro del programa de la Lic. en Ing. en Sistemas Inteligentes.	7
Secuencia Didáctica	8
Presentación	9
PRÁCTICA 1	10
INSTALANDO EL SOFTWARE WEKA	
Objetivo	10
Introducción	10
Desarrollo	11
Conclusión	14
Bibliografía	14
PRÁCTICA 2	15
CONOCIENDO EL SOFTWARE WEKA	
Objetivo	15
Introducción	15
Desarrollo	16
Conclusión	18
Bibliografía	18
PRÁCTICA 3	19
IMPLEMENTANDO ALGORITMOS DE CLASIFICACIÓN	
Objetivo	19
Introducción	19
Desarrollo	19
Conclusión	20
Bibliografía	20
PRÁCTICA 4	21
DEFINIENDO ALGORITMOS DE CLASIFICACIÓN	
Objetivo	21
Introducción	21
Desarrollo	22
Conclusión	23
Bibliografía	23
PRÁCTICA 5	24
ALGORITMOS DE CLUSTERING (CLUSTER WEKA)	
Objetivo	24
Introducción	24
Desarrollo	25
Conclusión	26
Bibliografía	26

PRÁCTICA 6	27
SELECCIÓN DE ATRIBUTOS (SELECT ATTRIBUTES)	
Objetivo	27
Introducción	27
Desarrollo	27
Conclusión	28
Bibliografía	29
PRÁCTICA 7	
ARCHIVOS ATTRIBUTE-RELATION FILE FORMAT (.ARFF)	30
Objetivo	30
Introducción	30
Desarrollo	31
Conclusión	32
Bibliografía	32
PRÁCTICA 8	33
CONOCIENDO EXPERIMENTER DE WEKA	
Objetivo	33
Introducción	33
Desarrollo	35
Conclusión	36
Bibliografía	36
PRÁCTICA 9	37
EXPERIMENTANDO EN MODO AVANZADO	
Objetivo	37
Introducción	37
Desarrollo	38
Conclusión	39
Bibliografía	39
PRÁCTICA 10	40
CONOCIENDO SIMPLE CLI DE WEKA	
Objetivo	40
Introducción	40
Desarrollo	41
Conclusión	41
Bibliografía	41

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

DIRECTORIO INSTITUCIONAL

Dr. en Edu. Alfredo Barrera Baca
RECTOR

M. en E.U. y R.

Marco Antonio Luma Pichardo

Secretario de Docencia

M. en C

Jannet Valero Vilchis

Secretaria de Rectoría

Dra. en Ed.

Sandra Chávez Marín

Secretaria de Extensión y Vinculación

M. en Dis.

Juan Miguel Reyes Viurquez

Secretario de Administración

M. en L.A.

María del Pilar Ampudia García

Secretaria de Cooperación Internacional

Dr. en C.S.

Luis Raúl Ortiz Ramírez

Abogado General

Lic. en Com.

Gastón Pedraza Muñoz

Director General de Comunicación
Universitaria

M. en D.F.

Jorge Rogelio Zenteno Domínguez

Encargado del Despacho de la Contraloría
Universitaria

M. en A. **José Francisco Mejía Carbajal**

Secretario Particular Adjunto del Rector

Dr. en C.I.

Carlos Eduardo Barrera Díaz

Secretario de Investigación y Estudios
Avanzados

Dr. en A.

José Édgar Miranda Ortiz

Secretario de Difusión Cultural

M. en E.

Javier González Martínez

Secretario de Finanzas

Dr. en C.C.

José Raymundo Marcial Romero

Secretario de Planeación y Desarrollo
Institucional

Dra. en Dis.

Mónica Marina Mondragón

Secretaría de Cultura Física y Deporte

M. en R. I.

Jorge Bernáldez García

Secretario Técnico de la Rectoría

M. en A. P.

Guadalupe Ofelia Santamaría González

Directora General de Centros Universitarios y
Unidades Académicas Profesionales

Lic. En Act.

Angelita Garduño Gómez

Secretaria particular del Rector

CENTRO UNIVERSITARIO UAEM NEZAHUALCÓYOTL

DIRECTORIO

Maestro en Derecho Juan Carlos Medina Huicochea	ENCARGADO DEL DESPACHO DE LA DIRECCIÓN
Maestro en Ciencias José Antonio Castillo Jiménez	Subdirector Académico
Licenciado en Economía Ramón Vital Hernández	Subdirector Administrativo
Doctora en Ciencias Sociales María Luisa Quintero Soto	Coordinadora de Investigación y Estudios Avanzados
Licenciado en Administración de Empresas Víctor Manuel Durán López	Coordinador de Planeación y Desarrollo Institucional
Maestro en Ciencias Cesar Lucio Gutiérrez Ruiz	Coordinador de la Licenciatura en Comercio Internacional
Maestro en S.F. Carlos Anaya Hernández	Coordinadora de la Licenciatura en Educación para la Salud
Doctor en Ingeniería de los Sistemas Ricardo Rico Molina	Coordinador de la Licenciatura en Ingeniería en Sistemas Inteligentes
Maestro en Ciencias Ricardo Pacheco Ruiz	Coordinador de la Licenciatura en Ingeniería en Transporte
Maestro en Ciencias de la Computación Erick Nicolás Cabrera Álvarez	Coordinador de la Licenciatura en Seguridad Ciudadana Mixta
Maestro en Administración José Ramon CS. Garcia Ibarra	Coordinador de la Licenciatura en Seguridad Ciudadana Presencial

Ubicación de la asignatura de Taller de Weka, dentro del programa de la Lic. en Ingeniería en Sistemas Inteligentes

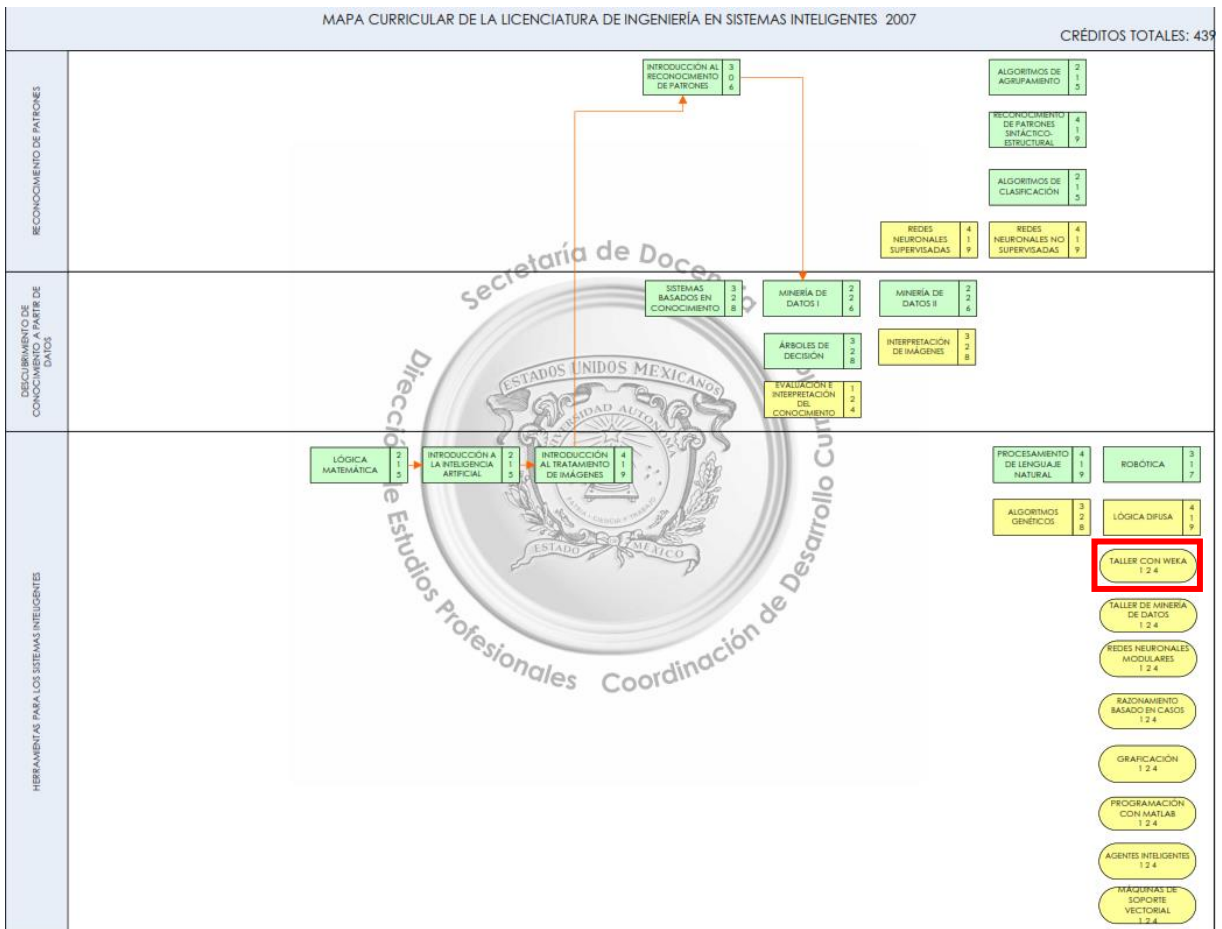


Figura 1. Mapa Curricular ISI-Taller de Weka

SECUENCIA DIDÀCTICA

PRÀCTICA 1. Instalando el Software Weka.

PRÀCTICA 2. Conociendo el Software Weka.

PRÀCTICA 3. Implementando Algoritmos de Clasificacion.

PRÀCTICA 4. Definiendo Algoritmos de Clasificacion.

PRÀCTICA 5. Algoritmos de Clustering (Cluster Weka).

PRÀCTICA 6. Selección de Atributos (Select Attributes).

PRÀCTICA 7. Archivos Attribute-Relation File Format (.arff).

PRÀCTICA 8. Conociendo Experimenter de Weka

PRÀCTICA 9. Experimentando en Modo Avanzado

PRÀCTICA 10. Conociendo Simple Cli de Weka

PRESENTACIÓN

El objetivo de este Manual de Prácticas de la Unidad de aprendizaje Taller de Weka es definir los procesos por los cuales se pueden llevar a cabo el análisis de datos con diferentes algoritmos.

Será una guía para la elaboración de cada una de las prácticas para poder manejar el software de Weka para los alumnos de decimo semestre del programa de Licenciatura de Ingeniería en Sistemas Inteligentes, siendo parte fundamental de las prácticas de dicha unidad, en el cual se organiza y coordina cada una de las actividades a realizar. Así mismo, se abordará desde una visión general de los principios, características y métodos del desarrollo en Weka siendo importante para el manejo de esta herramienta.

En este material el estudiante conocerá el marco conceptual y herramientas de cómputo que podrá utilizar con el propósito de responder a las demandas de integración de datos y análisis de información que demanda un proyecto del ámbito de la Ingeniería en Sistemas Inteligentes usando el software Weka.

PRÁCTICA 1

Instalando el Software Weka

Nombre del Alumno: _____

OBJETIVO

Conocer la forma de instalar el paquete de Weka y de esta manera dar un primer acercamiento al software y sus múltiples herramientas para el análisis de datos y poder dar un apoyo a la toma de decisiones.

INTRODUCCIÓN

La extracción de conocimiento a partir de fuentes masivas de datos. Para ello se emplean las denominadas técnicas de minería de datos, que son algoritmos capaces de obtener relaciones entre distintos atributos o conceptos para ayudar a la toma de decisiones. Además de las técnicas estadísticas se estudian las técnicas de Minería de Datos (Data Mining) basadas en técnicas de aprendizaje automático que se implementan en una herramienta de minería de datos de libre distribución: **WEKA**.

Esta herramienta permite, a partir de ficheros de texto en un formato determinado, utilizar distintos tipos de técnicas para extraer información. (García Herrero & Molina Lopez, 2012)

El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades.

Esta versión original se diseñó inicialmente como herramienta para analizar datos procedentes del dominio de la agricultura, pero la versión más reciente basada en Java (WEKA 3), que empezó a desarrollarse en 1997, se utiliza en muchas y muy diferentes áreas, en particular con finalidades docentes y de investigación.

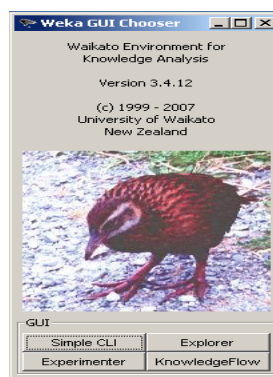


Ilustración 1: Weka GUI 1999

DESARROLLO

Para poder usar este software abriremos la página oficial de Weka (<https://www.cs.waikato.ac.nz/ml/Weka/downloading.html>) en el apartado de versiones estables utilizaremos la versión que no usa una máquina virtual (se instalará en la pc).

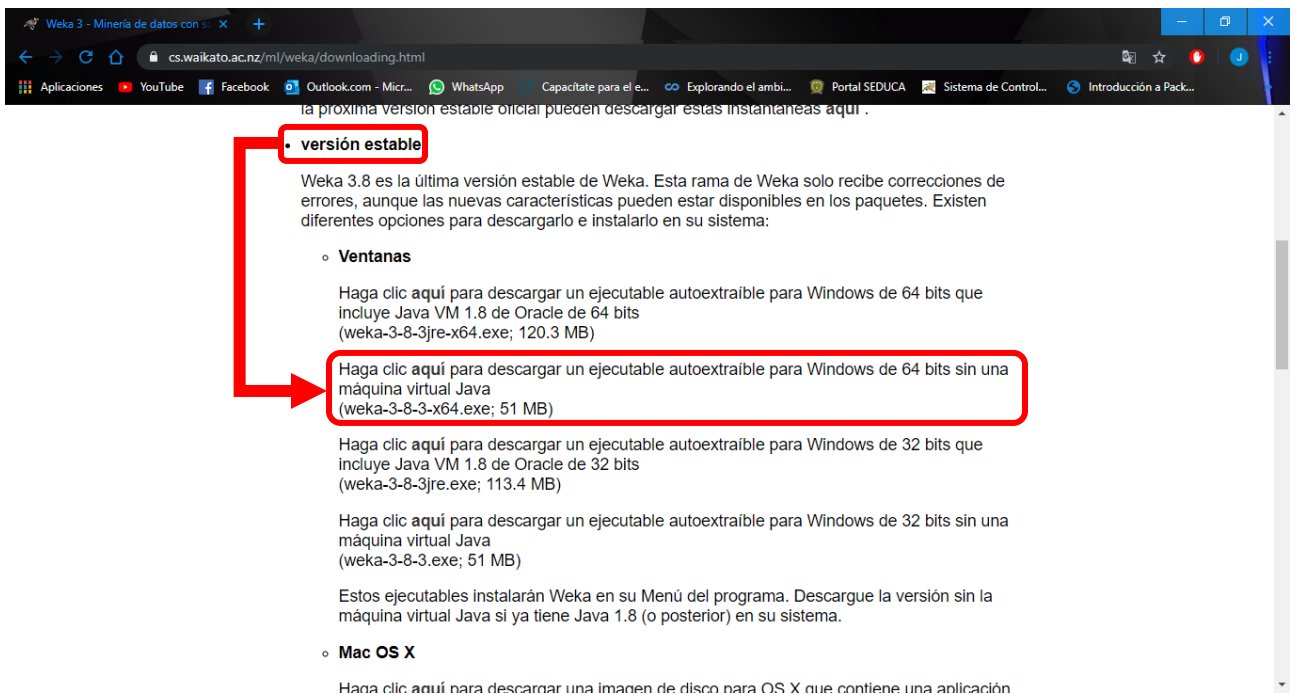


Ilustración 2: Pagina Weka Oficial de descarga 2019

Nota: En caso de usar Mac descargar del apartado Mac Os X.

Una vez hecho lo anterior nos redirigirá a una página en la cual después de 5seg iniciará la descarga automáticamente como vemos en la siguiente imagen:

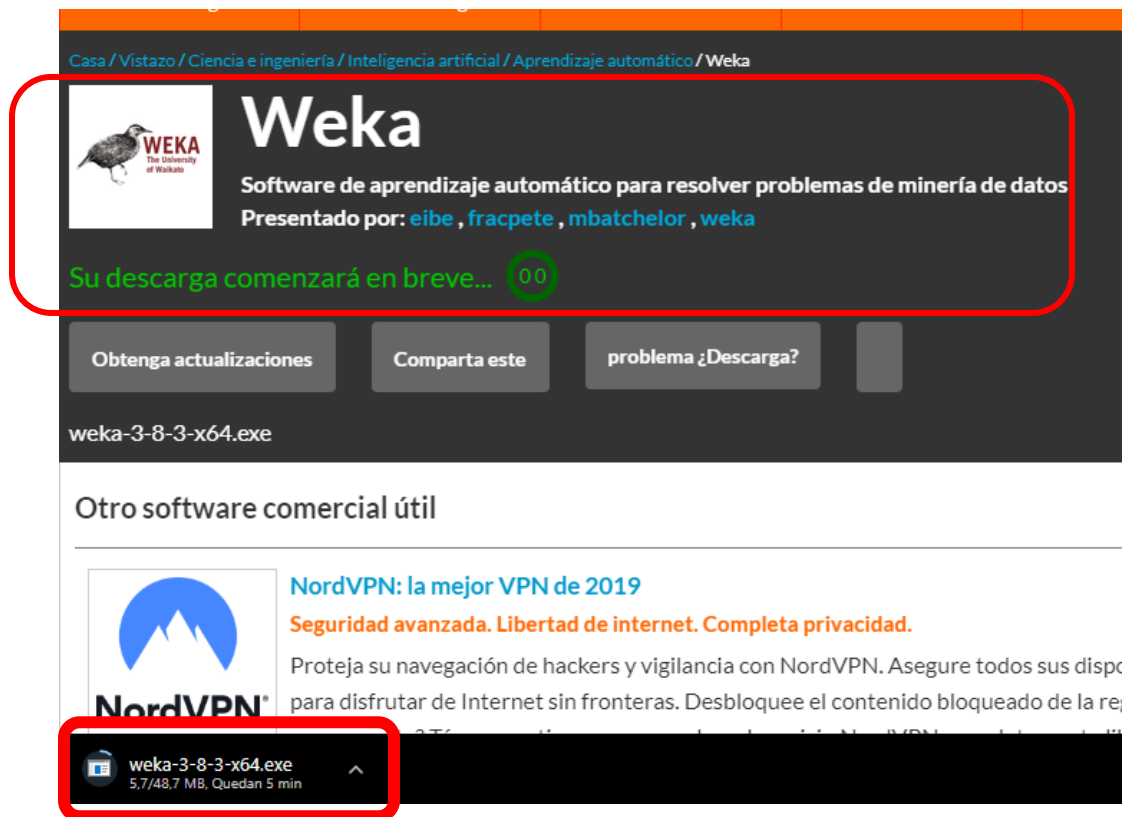


Ilustración 3: Descargando Weka

Cuando haya terminado la descarga se ejecutará como administrador el instalador y aparecerá la pantalla del Asistente de Instalación:

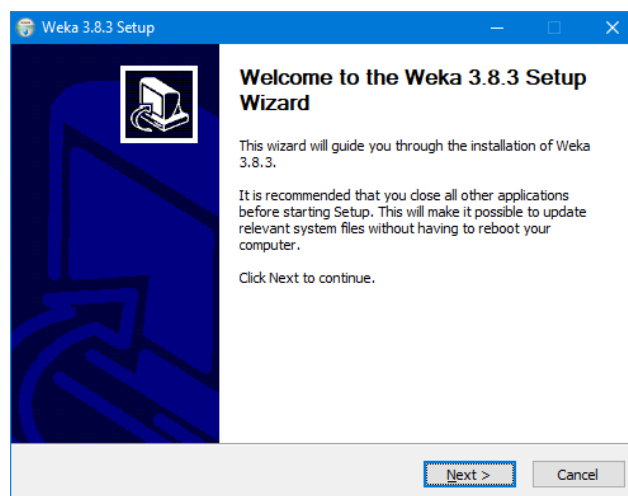


Ilustración 4: Ventana de Instalación de Weka

Daremos clic en Next, después aparecerá la pantalla en la cual nos redacta la licencia publica general y daremos clic en el botón “I Agree”.

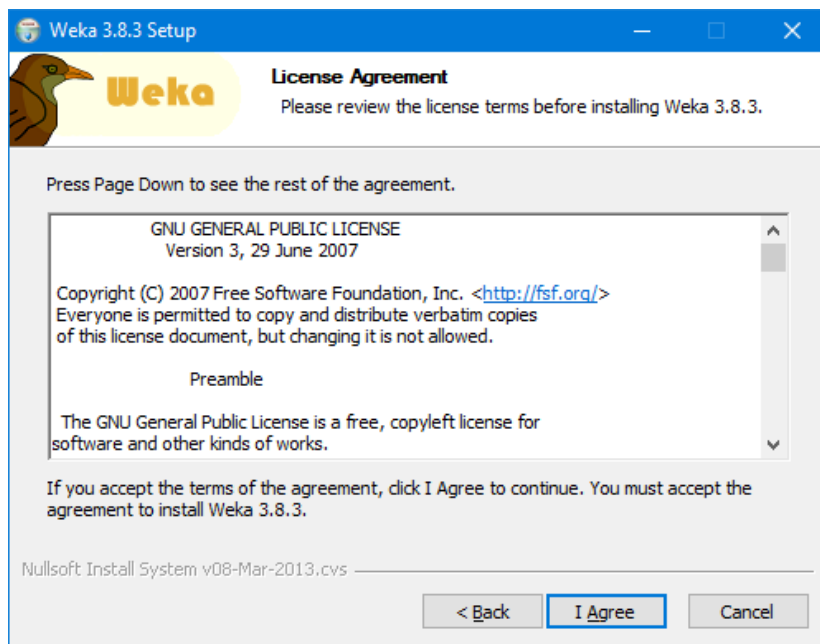


Ilustración 5: Acuerdos de Licencia Weka 2019

En esta pantalla dejamos todo por default dando clic en Next al igual que en las consecuentes.

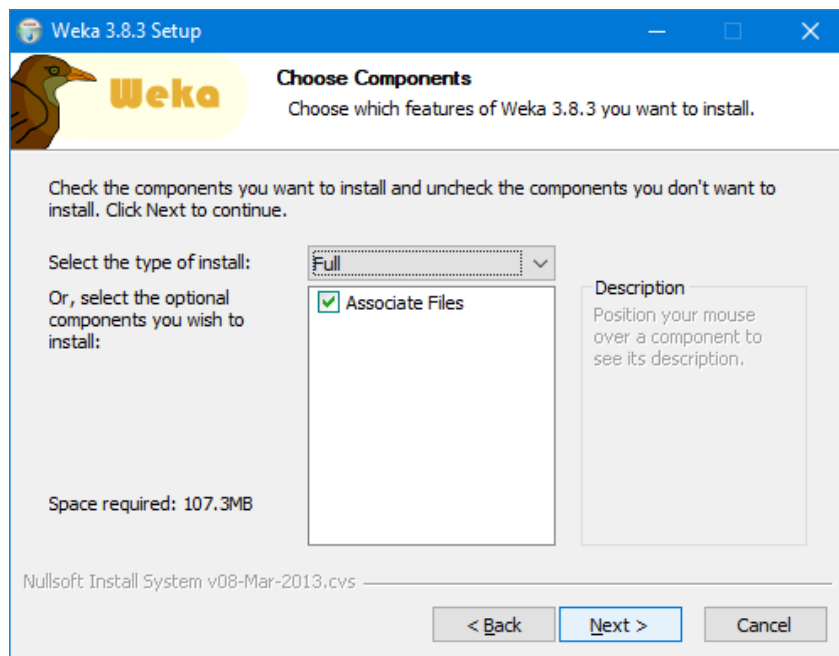


Ilustración 6: Eligiendo los Componentes

Una vez que haya terminado la instalación daremos clic en next y después en finish.

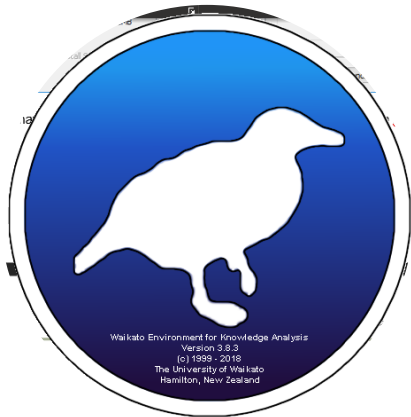


Ilustración 7: Bienvenida de Weka

Esta es la primera pantalla que nos mostrará una vez haya terminado de instalar el software. Posteriormente nos mostrará la siguiente pantalla:

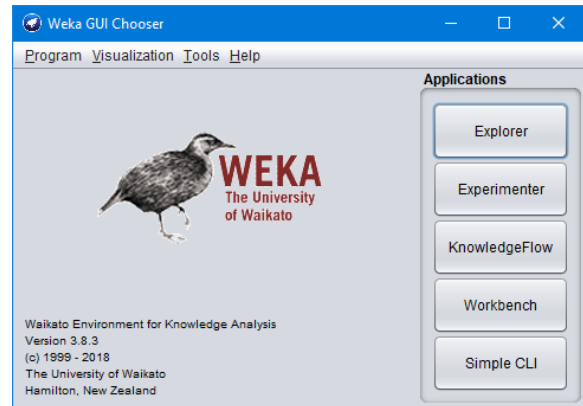


Ilustración 8: Ventana Principal

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica.

BIBLIOGRAFÍA

Garcia Herrero, Jesús; Molina López, José Manuel "Técnicas de Analisis de Datos (2012) Universidad Carlos III Madrid.

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

PRÁCTICA 2

Conociendo el Software Weka

Nombre del alumno: _____

OBJETIVO

Conocer cómo funciona el software Weka para poder hacer análisis sobre los metadatos de alguna base de datos y así mismo poder aplicar filtros (Algoritmos) para mejorar el análisis de los mismos.

INTRODUCCIÓN

El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. La versión original de Weka fue un front-end en TCL/TK para modelar algoritmos implementados en otros lenguajes de programación, más unas utilidades para preprocesamiento de datos desarrolladas en C para hacer experimentos de aprendizaje automático. Esta versión original se diseñó inicialmente como herramienta para analizar datos procedentes del dominio de la agricultura, pero la versión más reciente basada en Java (WEKA 3), que empezó a desarrollarse en 1997, se utiliza en muchas y muy diferentes áreas, en particular con finalidades docentes y de investigación. (García Morate)

Las características de Weka son:

- Está disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para preprocesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.

Weka soporta varias tareas estándar de minería de datos, especialmente, preprocesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Todas las técnicas de Weka se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos) (García Morate).

DESARROLLO

Por medio de la herramienta Weka precargue un repositorio de datos que viene como parte de los ejemplos a usar y su funcionamiento. Dando clic en el botón de “Open file” y después en la ventana selecciona la ruta “C:\Program Files\Weka-3-8\data” que es donde se instaló Weka y seleccionará el archivo con nombre “weather.nominal.arff”.

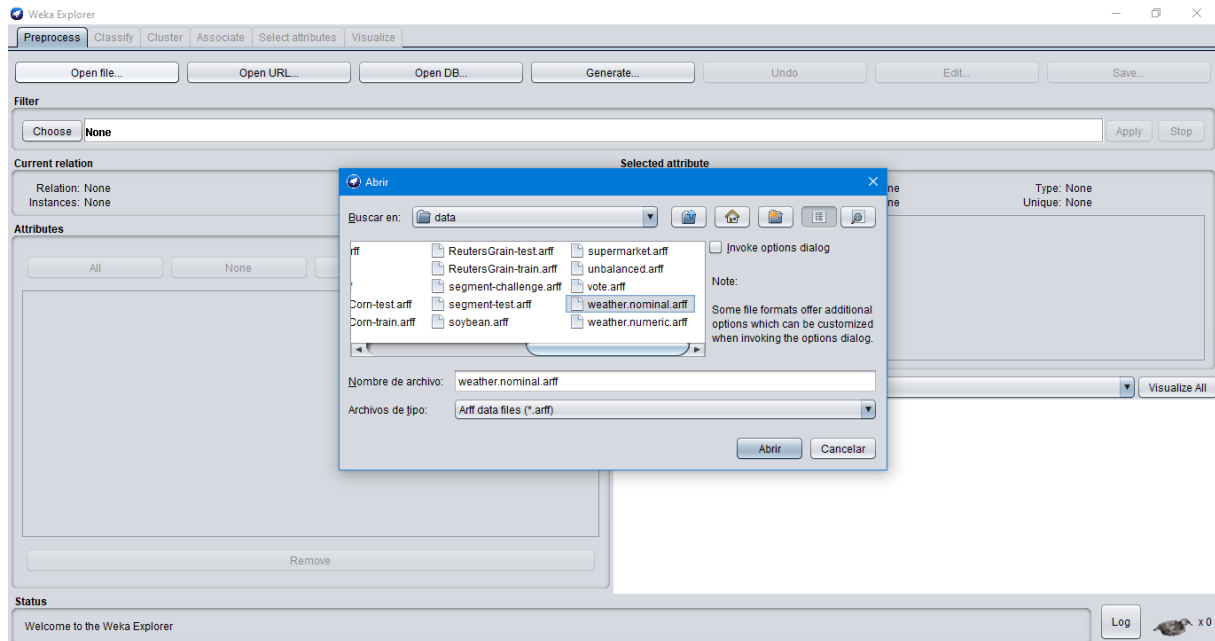


Ilustración 9: Carga de los Datos

El primer apartado que debe observar es el de Atributos (Attributes) en el cual aparecen todos los aspectos que representan a una instancia y que darán un resultado (una predicción).

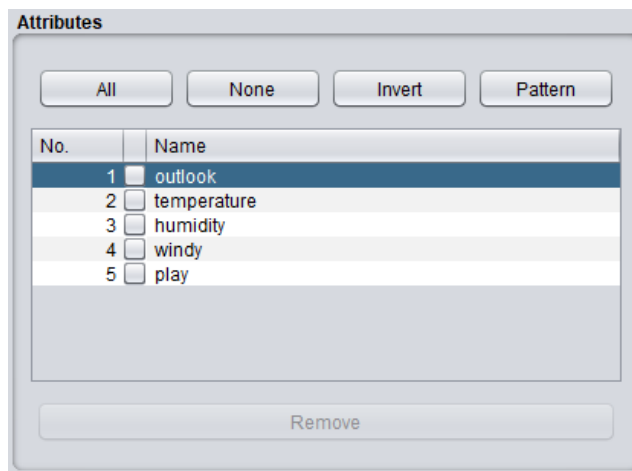


Ilustración 10: Apartado de Atributos de Weka

En caso de que requiera visualizar o modificar el set de datos (base de datos) hacer clic en el botón “Edit”.

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Ilustración 11: Ventana para editar los datos de entrada para Weka

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open UR... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** | Apply | Stop

Current relation: Relation: weather.symbolic | Instances: 14 | Attributes: 5 | Sum of weights: 14

Selected attribute: Name: outlook | Type: Nominal | Missing: 0 (0%) | Distinct: 3 | Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Attributes: All | None | Invert | Pattern

No.	Name
<input checked="" type="checkbox"/>	1 outlook
<input type="checkbox"/>	2 temperature
<input type="checkbox"/>	3 humidity
<input type="checkbox"/>	4 windy
<input type="checkbox"/>	5 play

Remove

Class: play (Nom) | Visualize All

Status: OK | Log | x 0

Ilustración 12: Explorador de Weka con datos cargados

Tomando en cuenta lo anterior y con el set de datos cargados visualice cual es la posibilidad de salir a jugar siendo que “temperatura=templado(mild)” o “humedad=normal”.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFIA

García Herrero, Jesús; Molina López, José Manuel “Técnicas de Análisis de Datos” (2012) Universidad Carlos III Madrid.

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

PRÁCTICA 3

Implementando algoritmos de clasificación

Nombre del Alumno: _____

OBJETIVO

Conocer algunos de los diferentes algoritmos que se encuentran disponibles en el software Weka para tener un análisis de datos estructurado

INTRODUCCIÓN

Algoritmos de clasificación supervisada. Los algoritmos de clasificación supervisada se utilizan en problemas en los cuales se conoce a priori el número de clases y los representantes de cada clase. Básicamente consiste en que, para clasificar automáticamente una nueva muestra, se tiene en cuenta la información que se pueda extraer de un conjunto de objetos disponibles divididos en clases y la decisión de una regla de clasificación o clasificador. (Riquelme C, Ruiz, & Gilbert, 2006)

Los algoritmos dedicados al problema de la clasificación supervisada operan usualmente sobre la información suministrada por un conjunto de muestras, patrones, ejemplos o prototipos de entrenamiento que son asumidos como representantes de las clases, y los mismos poseen una etiqueta de clase correcta. (Riquelme C, Ruiz, & Gilbert, 2006)

DESARROLLO

Haciendo uso del archivo "glass.arff" de la pestaña de Classify anote 5 diferentes algoritmos de clasificación aplicables a este archivo:

Después aplique el algoritmo de árboles (J48) al mismo archivo y visualice las instancias correctamente clasificadas y escriba el porcentaje de estas.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Garcia Herrero, Jesus; Molina Lopez, Jose Manuel "Tecnicas de Analisis de Datos" (2012) Universidad Carlos III Madrid.

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006.

PRÁCTICA 4

Definiendo Algoritmos de Clasificación

Nombre del Alumno: _____

OBJETIVO

Conocer cómo crear criterios de evaluación y modificación de algoritmos para el análisis de los metadatos.

INTRODUCCIÓN

Pulsando en la segunda pestaña (zona superior) del explorador entramos en el modo clasificación. En este modo podremos clasificar por varios métodos los datos ya cargados.

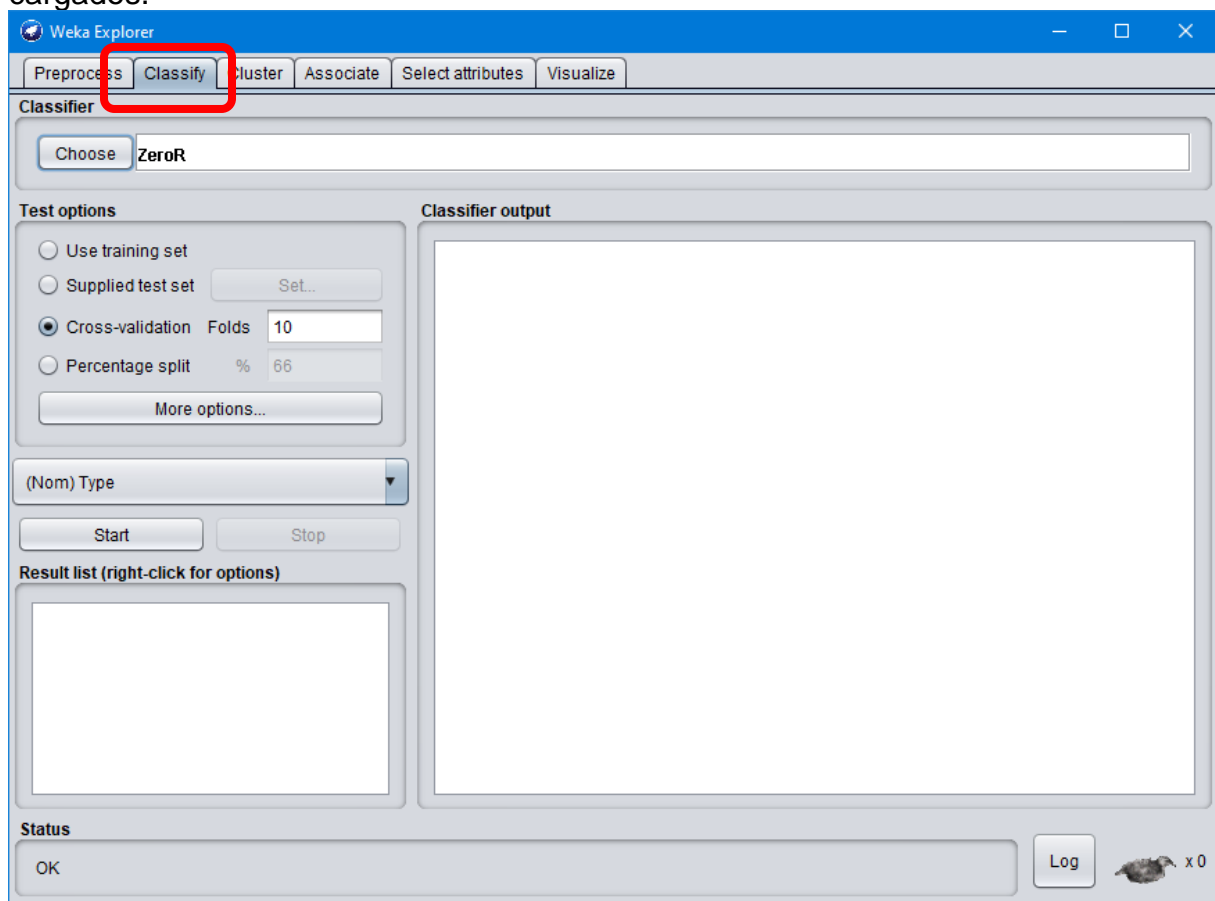


Ilustración 13: Ventana de clasificador de datos

Si queremos realizar una clasificación lo primero será elegir un clasificador y configurarlo a nuestro gusto, para ello pulsaremos sobre el botón Choose dentro del área Classifier. Una vez pulsado se desplegará un árbol que nos permitirá seleccionar el clasificador deseado. Una vez seleccionado aparecerá, en la etiqueta contigua al

botón Choose, el filtro seleccionado y los argumentos con los que se ejecutará. Esta información es muy útil si queremos utilizar el interfaz de consola ya que podremos configurar nuestro filtro con la interfaz y luego obtener el resultado apto para línea de mandato. Para poder acceder a las propiedades de cada clasificador deberemos hacer doble-clic sobre la etiqueta antes mencionada. Al darle aparecerá una nueva ventana con las propiedades junto a una breve explicación de este. Una vez elegido el clasificador y sus características el próximo paso es la configuración del modo de entrenamiento (Test Options). Weka proporciona 4 métodos de prueba:

1. **Supplied test set:** Marcando esta opción tendremos la oportunidad de seleccionar, pulsando el botón Set..., un fichero de datos con el que se probará el clasificador obtenido con el método de clasificación usado y los datos iniciales.
2. **Cross-Validation:** Pulsando el botón Cross-validation Weka realizará una validación cruzada estratificada del número de particiones dado (Folds). La validación cruzada consiste en: dado un número n se divide los datos en n partes y, por cada parte, se construye el clasificador con las $n-1$ partes restantes y se prueba con esa. Así por cada una de las n particiones.
3. **Percentage Split:** Se define un porcentaje con el que se construirá el clasificador y con la parte restante se probará
4. **Use Training set:** Con esta opción Weka entrenará el método con todos los datos disponibles y luego lo aplicará otra vez sobre los mismos.

Una vez definido el método de prueba también dispone de una opción "More option" el cual entre sus opciones nos da: Modelo de salida (Output Model), Salida de Estadísticas por clase(Output per-class stats), etc. (Garcia Morate)

DESARROLLO

Defina 3 algoritmos de árboles con las opciones que usted decida y aplique sobre el archivo glass.arff y reporte los resultados obtenidos en la matriz de confusión y las instancias correctamente clasificadas.

Algoritmo 1.- Nombre del Algoritmo: _____

Algoritmo 2.- Nombre del Algoritmo: _____

Algoritmo 3.- Nombre del Algoritmo: _____

Para finalizar deberá mostrar un solo árbol generado con cualquiera de los atributos antes realizados. Para ello en la zona inferior izquierda en la zona de resultados dará clic derecho y oprimirá la opción "Visualice tree" y aparecerá el árbol generado con los atributos usados.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

PRÁCTICA 5

Algoritmos de Clustering (Cluster Weka)

Nombre del alumno: _____

OBJETIVO

Conocer los diferentes algoritmos de Clustering (Agrupamiento) e identificar los óptimos al momento de analizar metadatos

INTRODUCCIÓN

Un algoritmo de agrupamiento (en inglés, clustering) es un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general distancia o similitud. La cercanía se define en términos de una determinada función de distancia, como la euclídea, aunque existen otras más robustas o que permiten extenderla a variables discretas. La medida más utilizada para medir la similitud entre los casos es la matriz de correlación entre los $n \times n$ casos. Sin embargo, también existen muchos algoritmos que se basan en la maximización de una propiedad estadística llamada verosimilitud. (Garcia Morate)

Generalmente, los vectores de un mismo grupo (o clústeres) comparten propiedades comunes. El conocimiento de los grupos puede permitir una descripción sintética de un conjunto de datos multidimensional complejo. De ahí su uso en minería de datos. Esta descripción sintética se consigue sustituyendo la descripción de todos los elementos de un grupo por la de un representante característico del mismo. (Garcia Morate)

Pulsando la tercera pestaña, llamada Clúster, en la parte superior de la ventana accedemos a la sección dedicada al clustering. El funcionamiento es muy similar al de clasificación: se elige un método de clustering, se selecciona las opciones pertinentes y con el botón "Start" empieza el funcionamiento. (Garcia Morate)

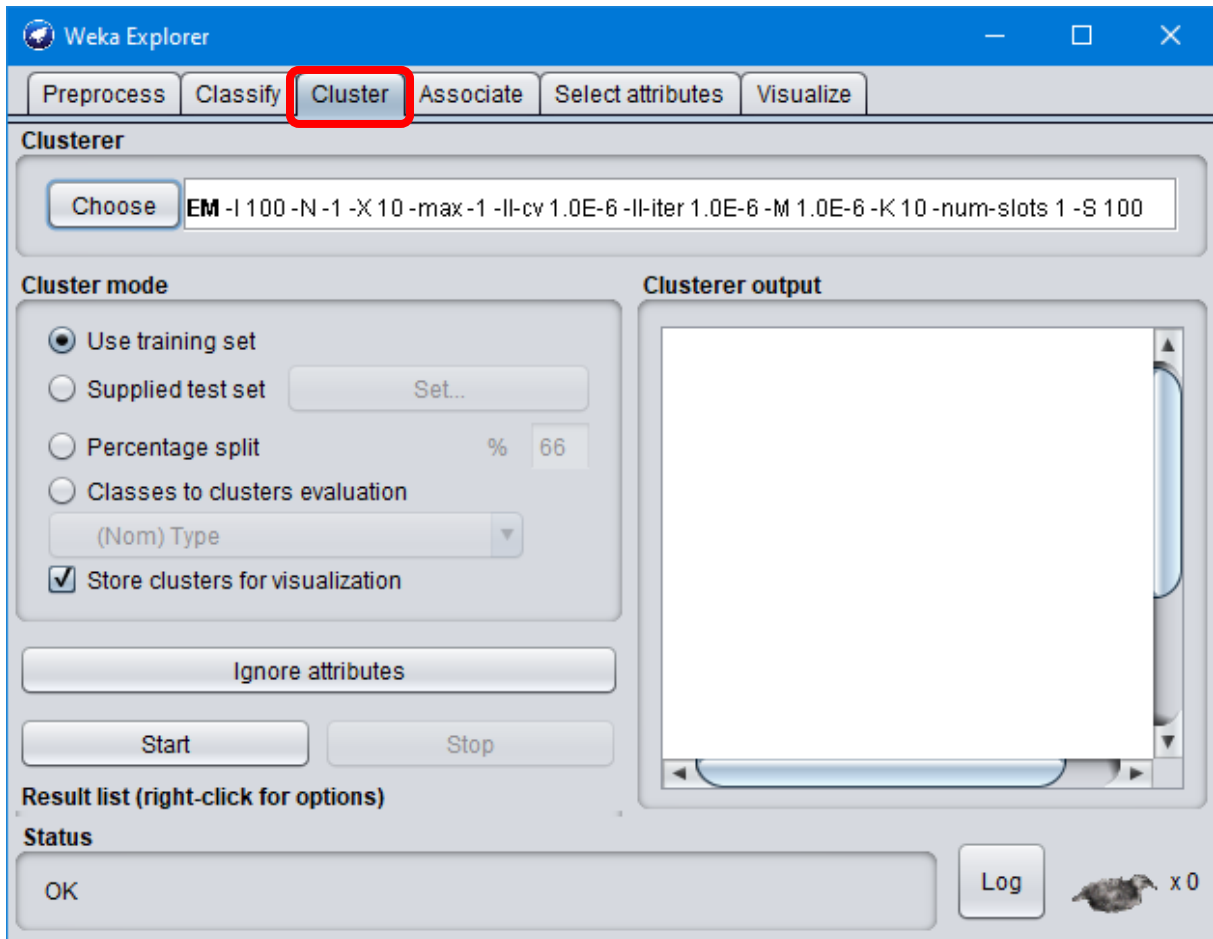


Ilustración 14: Ventana de Clúster de Weka

DESARROLLO

Mediante la opción de Clúster de Weka utilice el algoritmo “kmeans” con la opción “Percentage Split” en el archivo de “glass.arff” y reporte el clúster final de centroides en la siguiente tabla:

Tabla 1: Cluster final de Centroides

Attribute	Full data	Cluster
Rl		
Na		
Mg		
Al		
Si		
K		
Ca		
Ba		
Fe		

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

PRÁCTICA 6

Selección de Atributos (Select Attributes)

Nombre del alumno: _____

OBJETIVO

Conocer el funcionamiento de la pestaña de asociación para el análisis de datos.

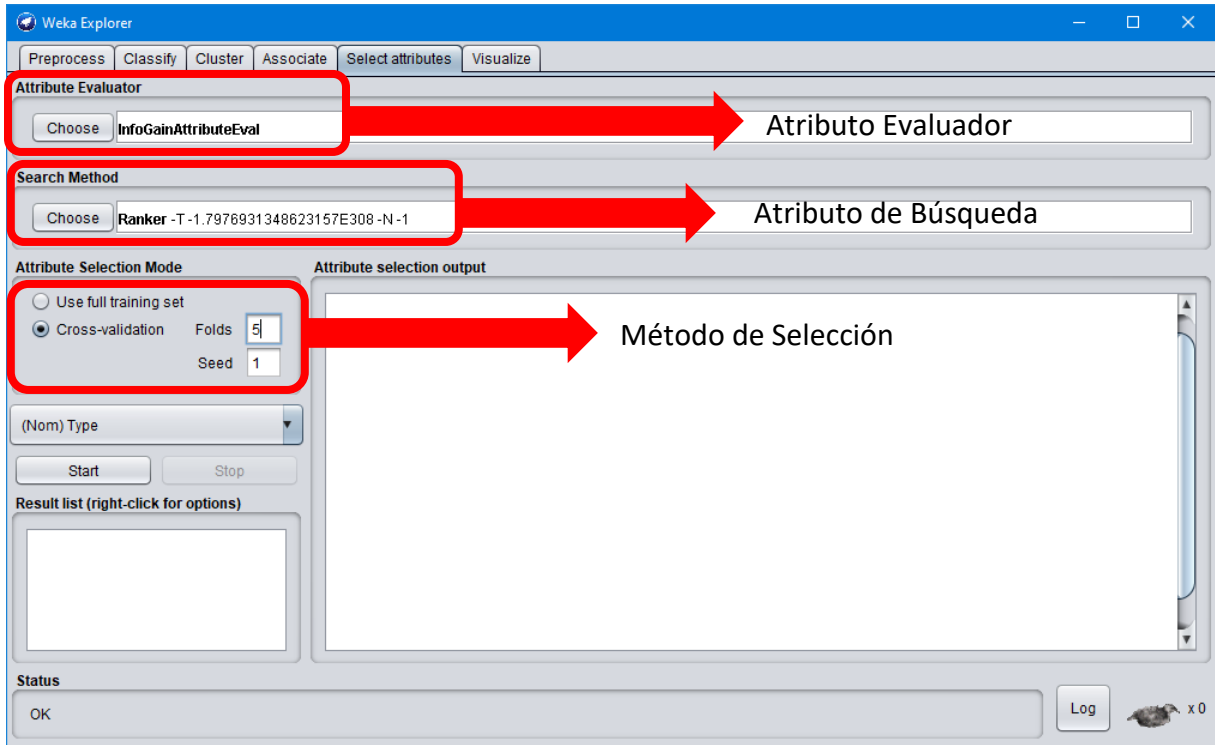
INTRODUCCIÓN

En esta parte trataremos la selección de atributos (eliminación de atributos redundantes e irrelevantes). Si hay un número excesivo de atributos, esto puede hacer que el modelo sea demasiado complejo y produzca un sobreajuste de valores dando como resultado valores no esperados. En Weka, la selección de atributos se puede hacer de varias maneras. La más directa es usando la 5ta pestaña "Select Attributes" con el cual deberemos seleccionar un método de búsqueda y un método de evaluación, existen 3 posibilidades:

1. Evaluación de Atributos
 - a. Método de Búsqueda = Ranker
 - b. Método de Evaluación = InfoGainAttributeEval
2. Evaluación de Conjuntos de Atributos: Filter
 - a. Método de Búsqueda = Greedy Stepwise
 - b. Método de Evaluación = CfsSubsetEval
3. Método Wrapper
 - a. Método de Búsqueda = Ranker
 - b. Método de Evaluación = InfoGainAttributeEval

DESARROLLO

Usando el método de Evaluación de Atributos con un crossvalidation de 5 hojas reporte el resultado obtenido



CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

Aler,Ricardo(2009) Tutorial de Weka 3.6.0

PRÁCTICA 7

Archivos Attribute-Relation File Format (.arff)

Nombre del alumno: _____

OBJETIVO

- Conocer cómo se encuentra estructurado un archivo arff
- Desarrollar un archivo arff básico

INTRODUCCIÓN

Nativamente Weka trabaja con un formato denominado arff, acrónimo de Attribute-Relation File Format. Este formato está compuesto por una estructura claramente diferenciada en tres partes: Cabecera, Declaración de Atributos y Sección de Datos. (García Morate)

En la cabecera se define el nombre de la relación que será de tipo String y en caso de contener algún espacio este deberá ser expresado con comillas. Su sintaxis es la siguiente:

@relation <nombre_de_la_relacion>

En la declaración de atributos se extiende todos los atributos que tendrá el archivo junto con el tipo de dato que maneja cada uno de ellos, y su sintaxis es la siguiente:

@attribute <nombre-del-atributo-1> <tipo>

@attribute <nombre-del-atributo-2> <tipo>

..@attribute < nombre-del-atributo-n> <tipo>

Los tipos de datos que puede manejar Weka son los siguientes:

- A) Numeric: Expresa números Reales
- B) Integer: Expresa únicamente números enteros
- C) Date: Este tipo de dato expresa fechas, pero para poder usarlo debe ir entrecomillado y así mismo maneja las etiquetas del siguiente formato:

“ddmmyyyy hh:mm:ss”.

En caso de no requerir algún campo no se define.

- D) String: Expresa cadenas de texto.
- E) Enumerado: El identificador de este tipo consiste en expresar entre llaves y separados por comas los posibles valores (cadenas de texto o caracteres) que puede tomar el atributo

@attribute nombre {valor1,valor2,...,valorn}

En la sección de datos se declara los datos que componen la relación separando entre comas los atributos y con saltos de línea las relaciones.

@data

4,3.2

Aunque éste es el modo “completo” es posible definir los datos de una forma abreviada (sparse data). Si tenemos una muestra en la que hay muchos datos que sean 0 podemos expresar los datos prescindiendo de los elementos que son nulos, rodeando cada una de las filas entre llaves y situando delante de cada uno de los datos el número de atributo

```
@data  
{1 4, 3 3}
```

En este caso hemos prescindido de los atributos 0 y 2 (como mínimo) y asignamos al atributo 1 el valor 4 y al atributo 3 el valor 3

En caso de no tener algún dato o tener un valor desconocido se expresará con el símbolo de interrogación (?) y para poder añadir un comentario en Weka se utilizará el símbolo del porcentaje (%) por cada símbolo de porcentaje se tomará una línea comentada, es decir, Weka no la procesará. (Garcia Morate)

Ejemplo de un programa .arff

```
%Archivo de prueba para mostrar como escribir un archivo para procesamiento
```

```
@relation prueba
```

```
@attribute Nombre STRING
```

```
@attribute Fecha_Nacimiento DATE “dd-MM-yyyy”
```

```
@attribute enfermedad_cronica {si,no}
```

```
@data
```

```
Antonio,“12-07-1995”,no
```

```
Ricardo,“25-04-1982”,si
```

```
Brenda,“02-11-1991”,si
```

```
Monica,“05-04-1985”,?
```

DESARROLLO

Con cualquier editor de texto de su preferencia genere un archivo con extensión .arff y posteriormente abra el archivo en el software Weka para hacer una clasificación con el algoritmo de árboles J48, escriba el código de su programa y reporte los resultados

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

Aler,Ricardo(2009) Tutorial de Weka 3.6.0

PRÁCTICA 8

Conociendo Experimenter de Weka

Nombre del alumno: _____

OBJETIVO

Conocer algunos de los procedimientos que se pueden llevar a cabo para el análisis de datos tomando uno o más métodos de clasificación.

INTRODUCCIÓN

El modo experimentador (Experimenter) es un modo muy útil para aplicar uno o varios métodos de clasificación sobre un gran conjunto de datos y, luego poder realizar contrastes estadísticos entre ellos y obtener otros índices estadísticos. (Garcia Morate)

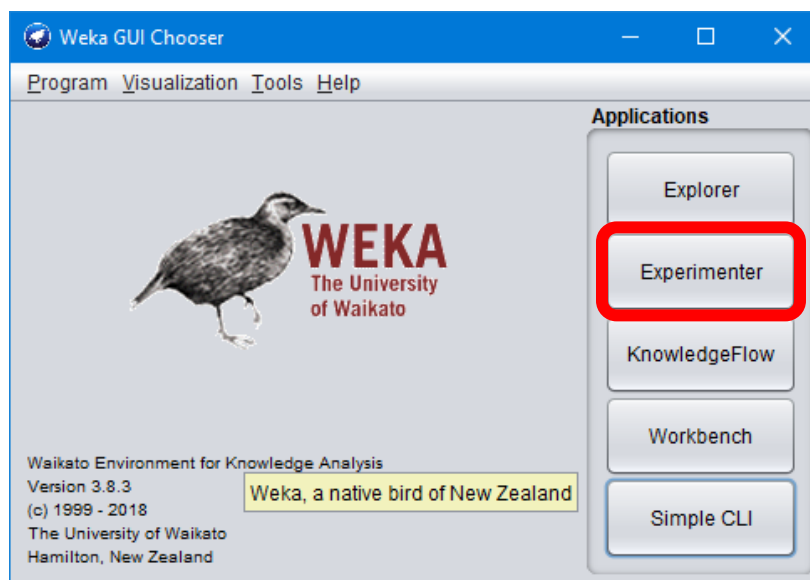


Ilustración 15: Ventana Principal Weka

Una vez abierto el modo experimentador obtendremos una ventana como la de la imagen, que corresponde a la sección Setup. Por defecto el modo experimentador está configurado en modo simple, no obstante, esto puede variarse a modo avanzado pulsando el botón circular acompañado con la etiqueta Advanced. Para empezar, se explicará el modo simple.

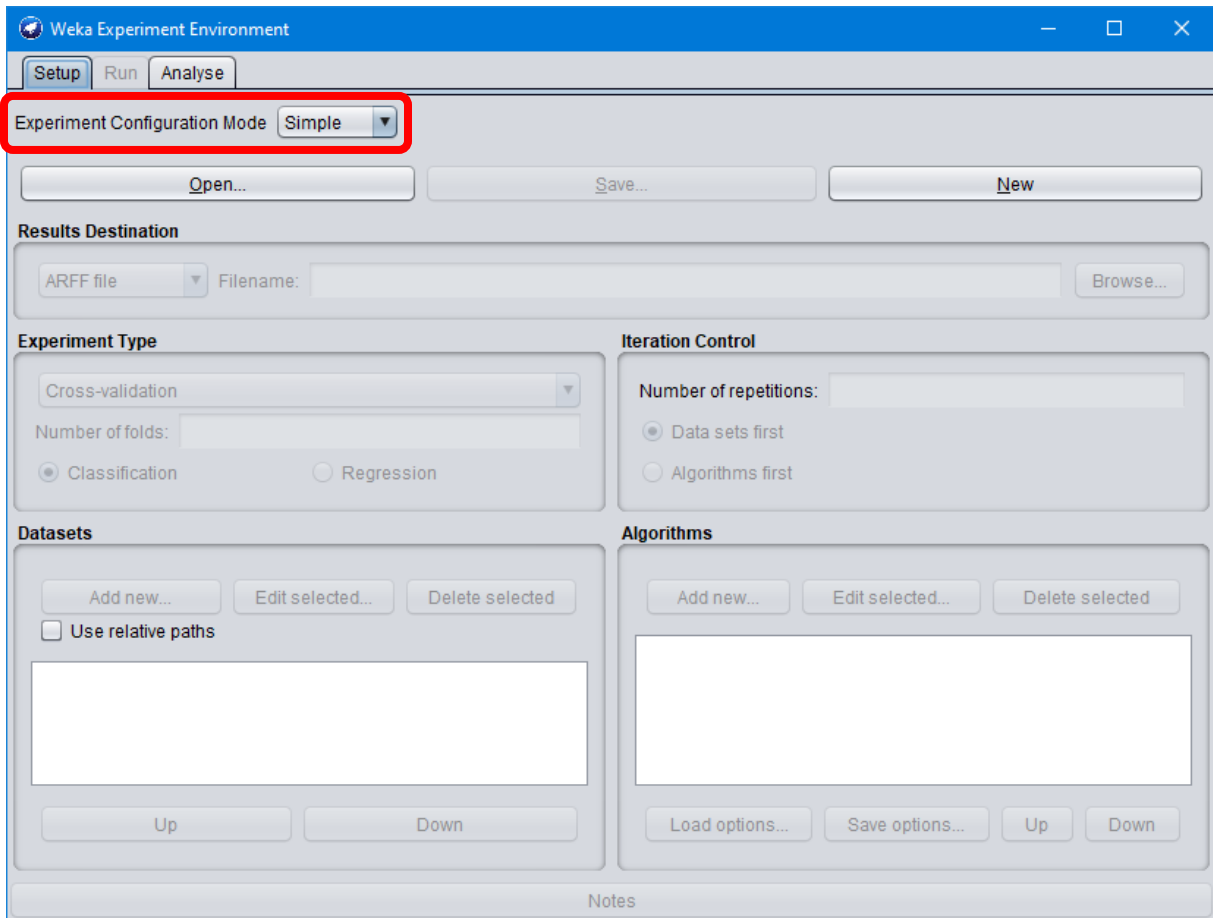


Ilustración 16: Ventana del Modo Experimento Simple de Weka

Lo primero a realizar es definir un fichero configuración que contendrá todos los ajustes, ficheros involucrados, notas, etc, pertenecientes a un experimento. Con el botón Open podemos abrir uno ya creado, pero también podremos crear uno nuevo con el botón “New”. Una vez creado podemos guardarlo (Weka por defecto no guarda nada), para lo que usaremos el botón “Save”. (Garcia Morate)

El siguiente paso es decidir dónde queremos almacenar los resultados, si es que queremos hacerlo. En caso de decantarnos por no hacerlo, aunque Weka lo permite, luego no se podrán ver los resultados obtenidos. Podemos archivar el resultado del experimento de tres formas distintas, en un fichero arff, en un fichero CSV y en una base de datos. Para ello, simplemente debemos seleccionar la opción que corresponda dentro del cuadro Results Destination y a continuación elegir la ruta lógica dónde guardar dicho resultado. (Garcia Morate)

Lo siguiente es definir el tipo de validación que tendrá el experimento, al igual que en el modo Explorer, tenemos tres tipos de validación: validación-cruzada estratificada, entrenamiento con un porcentaje de la población tomando ese porcentaje de forma

aleatoria y entrenamiento con un porcentaje de la población tomando el porcentaje de forma ordenada.

Una vez completados los anteriores pasos, deberemos indicar a Weka qué archivos de datos queremos que formen parte de nuestro experimento. Para ello en el cuadro Datasets tenemos un botón llamado Add new. . . que nos permite especificar un archivo de datos a añadir, o bien, seleccionando un directorio, nos añadirá todos los archivos que contengan ese directorio de forma recursiva. Junto a ese botón se encuentra la opción Use relative paths, que se utiliza para indicar a Weka que almacene las referencias a los ficheros como referencias relativas no indirectas. Puede ser útil si deseamos repetir el experimento en otro ordenador distinto con una estructura de ficheros distinta.

En el cuadro denominado Iteration control definiremos el número de repeticiones de nuestro experimento, especificando si queremos que se realicen primero los archivos de datos o los algoritmos. Debajo de este cuadro se encuentra la sección de algoritmos en el que con el botón Add new. . . podremos añadir los algoritmos que queramos. Si queremos borrar alguno lo seleccionamos y pulsamos el botón Delete selected. (García Morate)

Si se desea se pueden añadir comentarios y notas al experimento; para ello, Weka dispone de un cuadro de texto en la parte inferior de la ventana.

Con esto ya estaría configurado un experimento en modo simple.

DESARROLLO

Mediante el modo de experimentación utilice de los archivos existentes que proporciona Weka un archivo arff y un archivo csv con 2 algoritmos que usted elija. Reporte los resultados obtenidos. NOTA: Para poder visualizar los resultados vaya a la pestaña RUN en el botón Start.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Riquelme, José C.; Ruiz, Roberto; Gilbert, Karina. Minería de Datos: Conceptos y Tendencias. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 10, núm. 29, primavera 2006

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

Aler,Ricardo(2009) Tutorial de Weka 3.6.0

PRÁCTICA 9

Experimentando en modo Avanzado

Nombre del alumno: _____

OBJETIVO

Crear archivos de salida de experimentos en modo avanzado para poder aplicar diferentes algoritmos en conjunto a diferentes archivos unificados

INTRODUCCIÓN

En la configuración avanzada del modo experimento la principal diferencia es que el modo está orientado a realizar tareas específicas más concretas que un experimento normal, y con una cierta funcionalidad más explícita y concreta al usuario.

El cuadro “Destination” especifica dónde guardará Weka el resultado del experimento. Su funcionalidad es la misma que la del modo simple, pero no su funcionamiento. En este caso deberemos pulsar “Choose” y elegir en el árbol el tipo de método a utilizar para exportar los resultados (Base de datos, fichero arff o fichero csv).

Una vez elegido pulsando sobre la etiqueta que indica el método usado, se abrirá una ventana desplegable en donde configuraremos dicho método. Aunque pueda parecer una complicación el cambio en la interfaz para realizar esta operación, ésta segunda permite la adición de nuevos métodos sin variar la interfaz original.

El cambio más significativo entre la interfaz simple y avanzada, la veremos en el “*Result Generator*”. En este nuevo modo es necesario seleccionar con el botón “Choose” que método generador de resultados utilizaremos y seguidamente configurarle. Para esto pulsamos doble clic en la etiqueta que identifica el método elegido.

Weka admite 5 diferentes tipos de Generador de Resultados los cuales son:

- **CrossValidationResultProducer**: Este método genera resultados fruto de una validación cruzada, las opciones que maneja son:
 - **Numfolds**: número de particiones para la validación cruzada
 - **outputFile**: fichero donde queremos guardar cada una de las particiones que se realizarán, se guarda con formato .zip
 - **rawOutput**: guarda los resultados “tal cual”, es decir, sin ningún formato.
 - **splitEvaluator**: el método que seleccionemos para evaluar cada una de las particiones de la validación cruzada.
- **AveragingResultProducer**: Toma los resultados de un método generador y calcula los promedios de los resultados. Es útil, por ejemplo, para realizar una validación cruzada de muchas particiones por separado y hallar los promedios. Tiene las siguientes opciones:
 - **calculateStdDevs**: Seleccionado calcula las desviaciones estándar.

- **expectedResultsPerAverage:** Se elige el número de resultados que se esperan, para dividir la suma de todos los resultados entre este número.
- **keyFieldName:** En esta opción se selecciona el campo que será distintivo y único de cada repetición.
- **LearningRateResultProducer:** Llama a un método generador de resultados para ir repitiendo el experimento variando el tamaño del conjunto de datos. Habitualmente se usa con un AveragingResultProducer y CrossValidationResultProducer para generar curvas de aprendizaje. Sus opciones son:
 - **Lowersize:** Selecciona el número de instancias para empezar. Si es 0 se toma como mínimo el tamaño de stepSize.
 - **resultProducer:** Selecciona el método generador de resultados. Habitualmente para hacer curvas de aprendizaje se usa el AveragingResultProducer.
 - **Setzise:** Número de instancias a añadir por iteración.
 - **upperSize:** Selecciona el número máximo de instancias en el conjunto de datos. Si se elige -1 el límite se marca en el número total de instancias.
- **RandomSplitResultProducer:** Genera un conjunto de entrenamiento y de prueba aleatorio para un método de clasificación dado.
- **DatabaseResultProducer:** De una base de datos toma los resultados que coinciden con los obtenidos con un método generador de resultados (cualquiera de los anteriores). Si existen datos que no aparezcan en la base de datos se calculan.
 - **cacheKeyName:** Selecciona el nombre del campo de la base de datos a utilizar para formar una caché de la que coger los datos.
 - **databaseURL:** Dirección de la base de datos.
 - **Resultproducer:** Método generador de resultados a utilizar.
 - **Username:** Nombre de usuario para acceder a la base de datos.
 - **Password:** Contraseña para acceder a la base de datos.

DESARROLLO

Realice un archivo con extensión.arff, úselo como dataset con el modo avanzado de Weka en la pestaña Experiment. Reporte los resultados obtenidos usando cualquier algoritmo que desee.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Acosta Franco, Javier. "Aplicación de los Sistemas Clasificadores tradicionales al análisis de datos. Adquisición automática de reglas". Proyecto Fin de Carrera, Universidad Carlos III de Madrid, 2002.

Aler,Ricardo(2009) Tutorial de Weka 3.6.0

PRÁCTICA 10

Conociendo Simple Cli de Weka

Nombre del alumno: _____

OBJETIVO

Conocer el entorno de consola de Weka

INTRODUCCIÓN

Simple CLI es una abreviación de Simple Client. Esta interfaz proporciona una consola para poder introducir mandatos. A pesar de ser en apariencia muy simple es extremadamente potente porque permite realizar cualquier operación soportada por Weka de forma directa; no obstante, es muy complicada de manejar ya que es necesario un conocimiento completo de la aplicación. Su utilidad es pequeña desde que se fue recubriendo Weka con interfaces. Actualmente ya prácticamente sólo es útil como una herramienta de ayuda a la fase de pruebas.

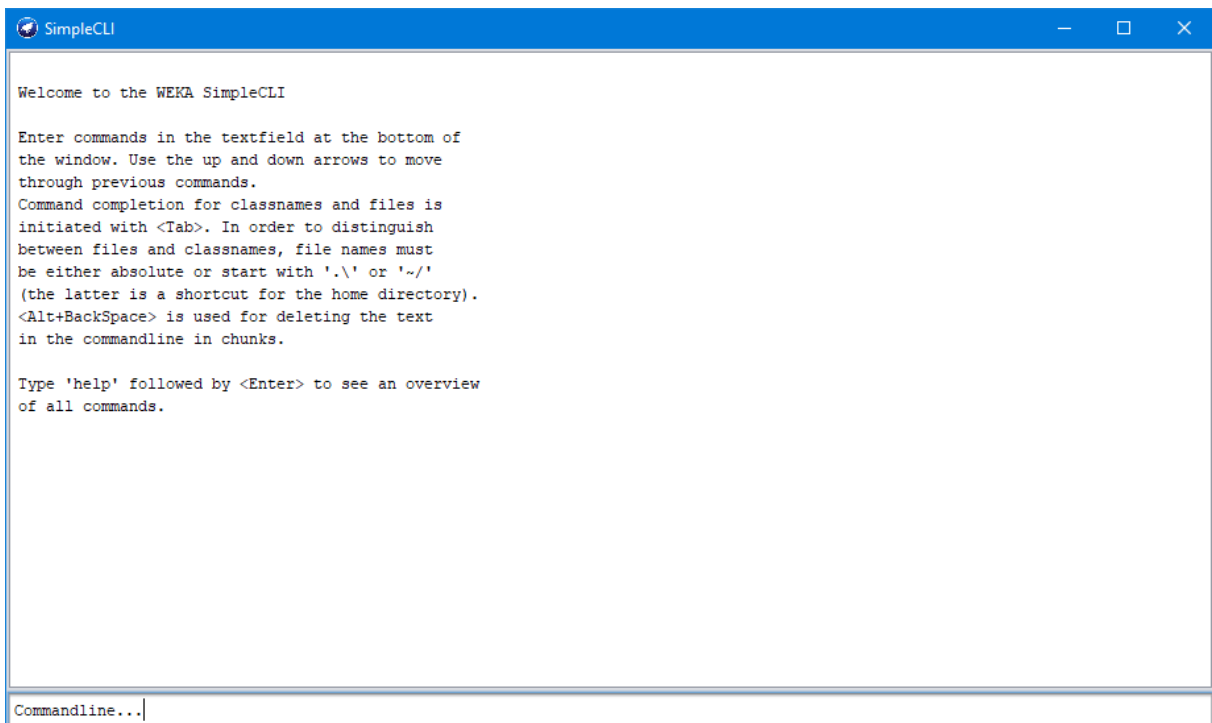


Ilustración 17: Ventana de Consola (Simple Cli)

Algunos de los comandos soportados por esta consola son:

Java <nombre-de-la-clase> <args>: Ejecuta el método “main” de la clase dada con los argumentos especificados al ejecutar este mandato (en el caso de que realmente se haya proporcionado alguno). Cada mandato es atendido en un hilo independiente por lo que es posible ejecutar varias tareas concurrentemente.

break: Detiene la tarea en ejecución actual.

Kill: Mata la tarea actual. En caso de que el comando break no funcione este terminará el proceso sin más.

Exit: Sale de la aplicación

Help <mandato>: proporciona una breve descripción del mandato.

DESARROLLO

Realiza un ejercicio donde ejecutes los comandos soportados por la consola, para que veas como funciona y cuales es su sintaxis.

CONCLUSIONES

Anote de manera breve las principales conclusiones obtenidas al término de esta práctica

BIBLIOGRAFÍA

Rousseeuw, P.J.; Kaufman, L. (1990). Finding Groups in Data: An Introduction to Clúster Analysis. Wiley.

Aler,Ricardo(2009) Tutorial de Weka 3.6.0

