
**Alumno:**

Ricardo Jara.

**Materia:**

SE

**Ciclo:**

9no

**Fecha:**

24/07/2020

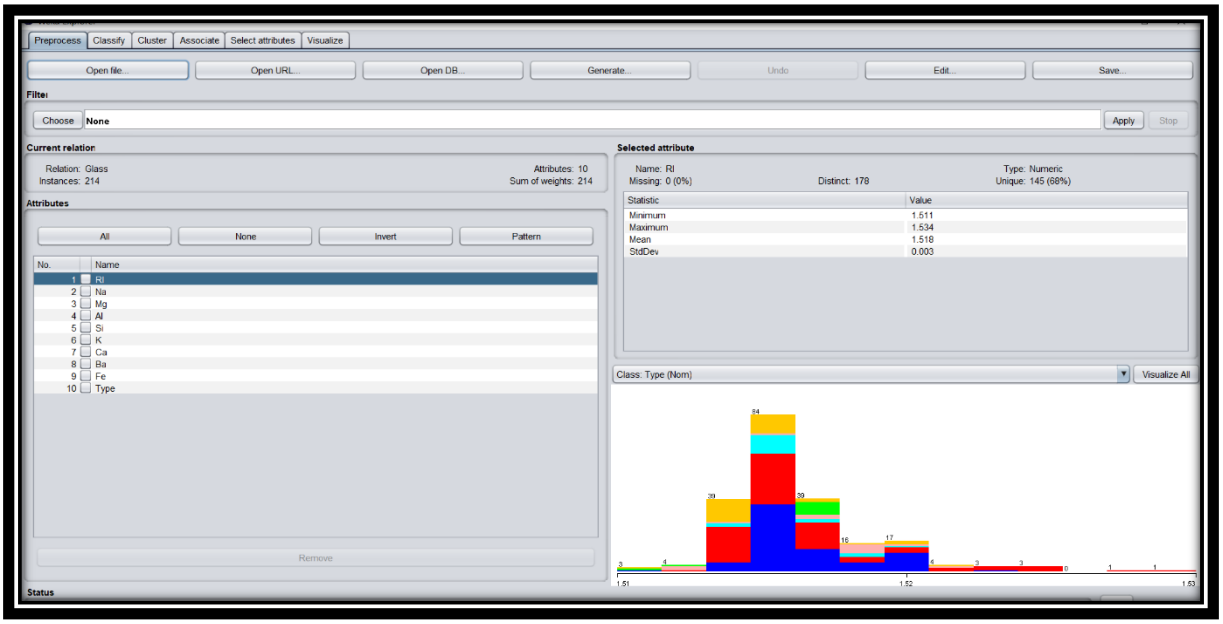
## Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato)

Entorno para análisis del conocimiento de la Universidad de Waikato, se denomina a sí mismo un conjunto de Librerías para tareas de minería de datos. Las librerías pueden ser llamadas desde la interfaz de weka o desde tus propias clases Java. Weka contiene herramientas para diferentes tareas básicas:

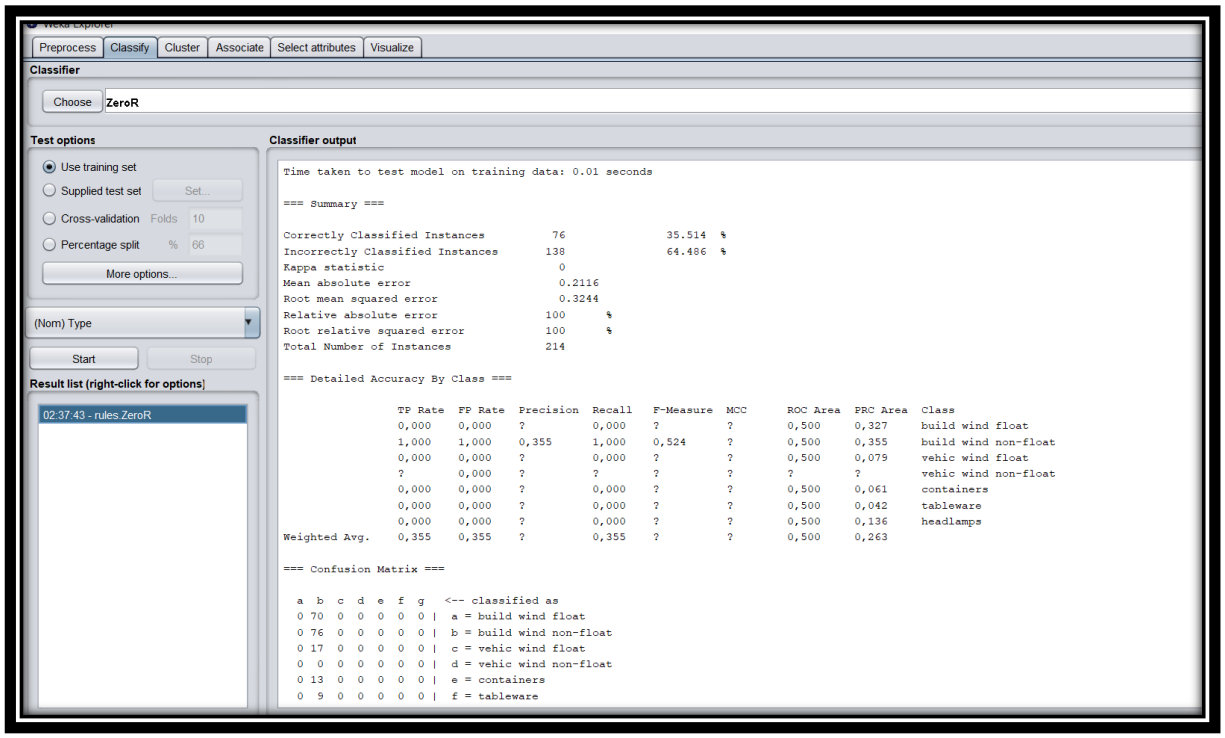
**Preprocess:** Multitud de herramientas para el preprocesamiento de los datos (como por ejemplo discretización de variables).
**Classify:** Algoritmos de clasificación, distribuidos por paquetes, como por ejemplo ID3 o C4.5
**Cluster:** Diferentes algoritmos de segmentación como el simple k-means.
**Associate:** Algoritmos para encontrar relaciones de asociación entre variables (Apriori entre otros).
**Select atributtes:** Aquí, una vez cargados los datos, Weka es capaz de buscar por nosotros las mejores variables del modelo.
**Visualize:** Herramienta de visualización de datos en los ejes cartesianos, con muchas posibilidades

A. Realizamos la carga de datos, en weka, para esto se eligió el dataset, **"glass.arff".**

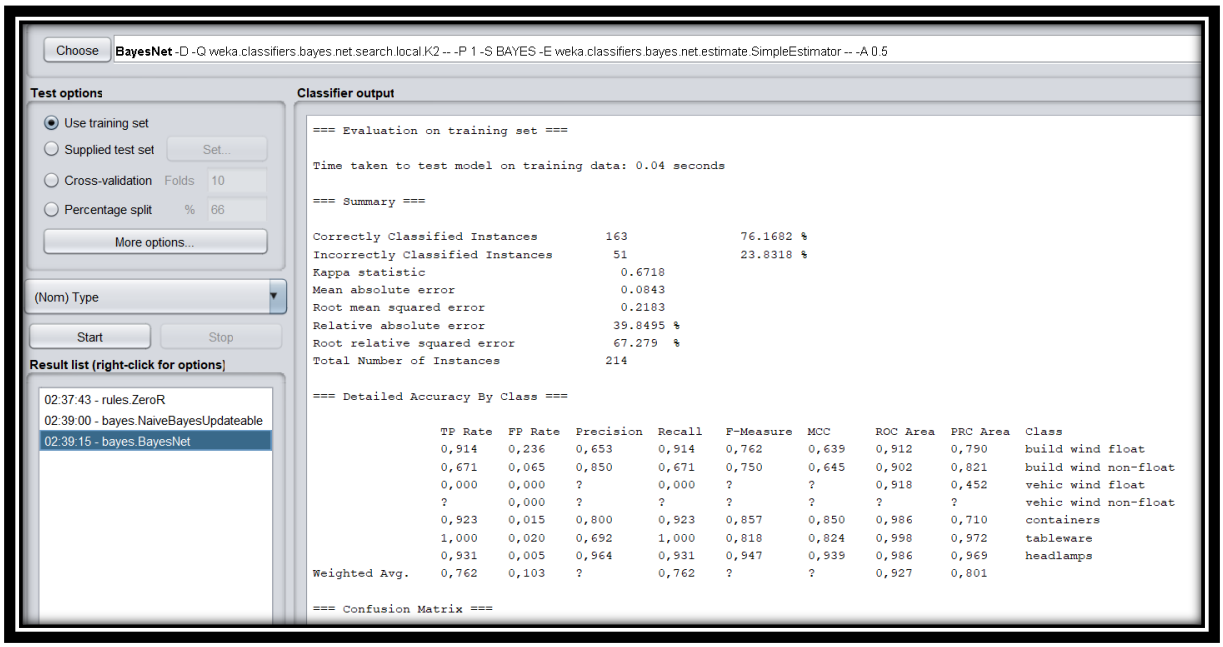
Vina realizó una prueba de comparación de su sistema basado en reglas, BEAGLE, el algoritmo vecino más cercano y análisis discriminante. BEAGLE es de un producto disponible a través de VRS Consulting, Inc Al determinar si el vidrio era un tipo de vidrio "flotante" o no, se obtuvieron los siguientes resultados (# respuestas incorrectas): Tipo de muestra Beagle NN DA De ventanas que fueron procesadas flotantemente (87) 10 12 21 De ventanas que no eran: (76) 19 16 22 El estudio de clasificación de tipos de vidrio fue motivado por investigación criminológica. En la escena del crimen, el cristal se fue El% se puede usar como evidencia ... ¡si se identifica correctamente! 5.
**Número de instancias: 214** **Título: Base de datos de identificación de vidrio**



1. Analizamos los datos mediante reglas, existen distintas opciones para poder realizar, la clasificación.




Se puede generar un test aplicando la regla de Bayes, clasificatoria donde nos dice los porcentajes de datos analizados, nos ubica en un clúster y nos arroja la cantidad de predicciones correctas e incorrectas, de un de 214.

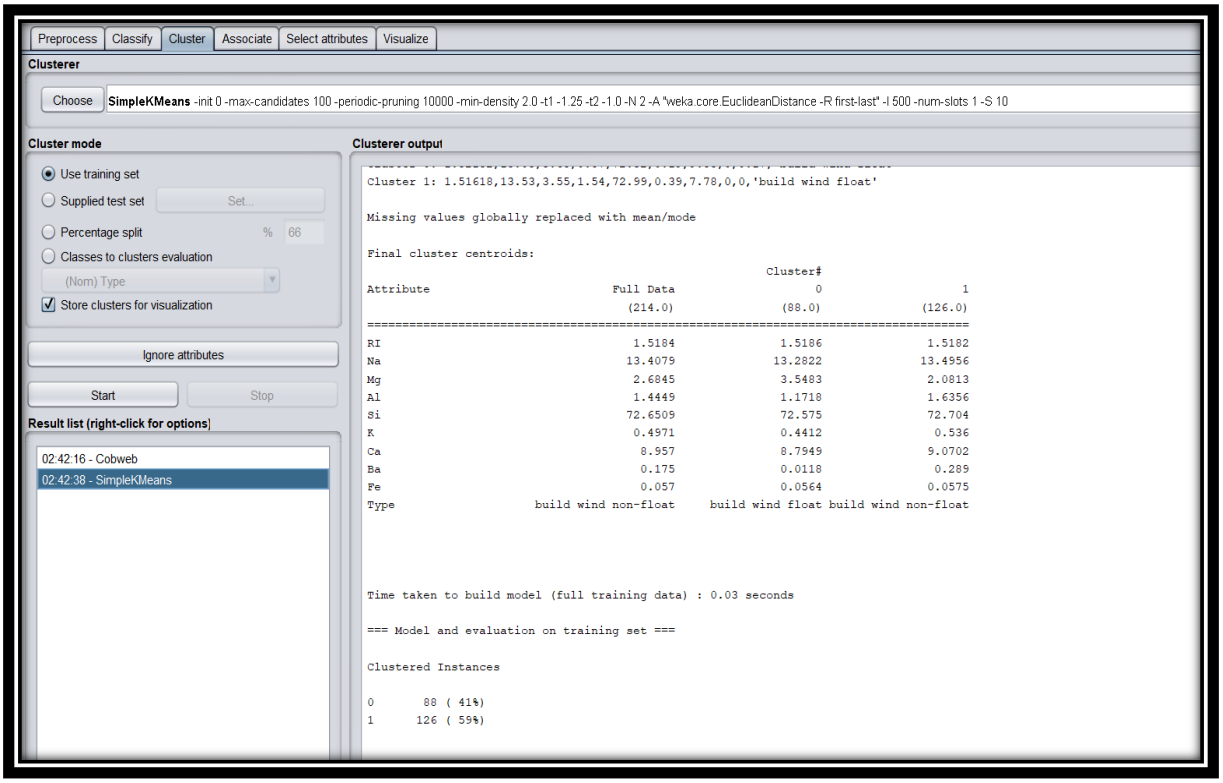


B. Aquí nos indica que realizo un total de 9 iteraciones, donde nos clasifica el vendedor, los porcentajes de venta y las marcas que han sido vendidas, este clasifica en dos clústeres.

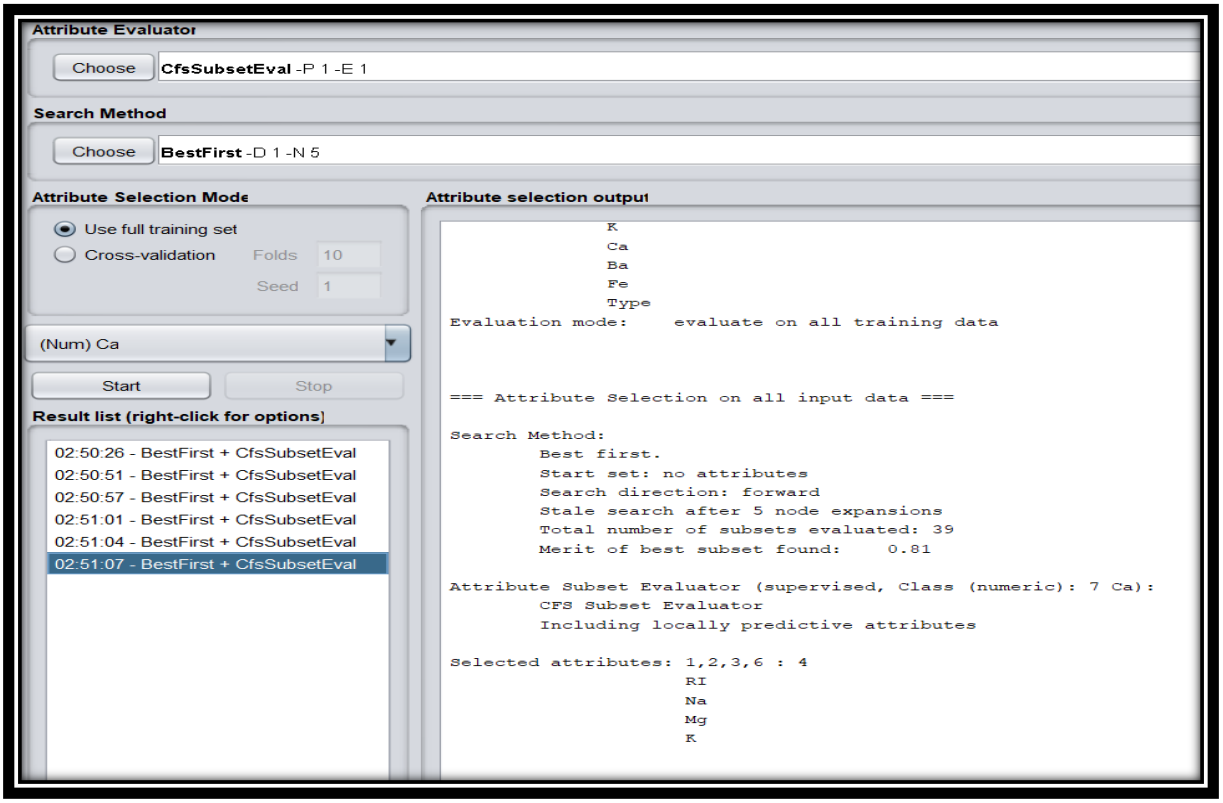
```
Number of iterations: 9
Within cluster sum of squared errors: 118.20374073549189

Initial starting points (random):

Cluster 0: 1.52152,13.05,3.65,0.87,72.32,0.19,9.85,0,0.17,'build wind float'
Cluster 1: 1.51618,13.53,3.55,1.54,72.99,0.39,7.78,0,0,'build wind float'

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                Cluster#
Attribute                    Full Data                0                1
                               (214.0)             (88.0)           (126.0)
```

C. Se puede ver en el grafico a continuación como clasifica en dos cluster lo que es valido para construir vidrio y el material que posiblemente no fuese útil para su uso.

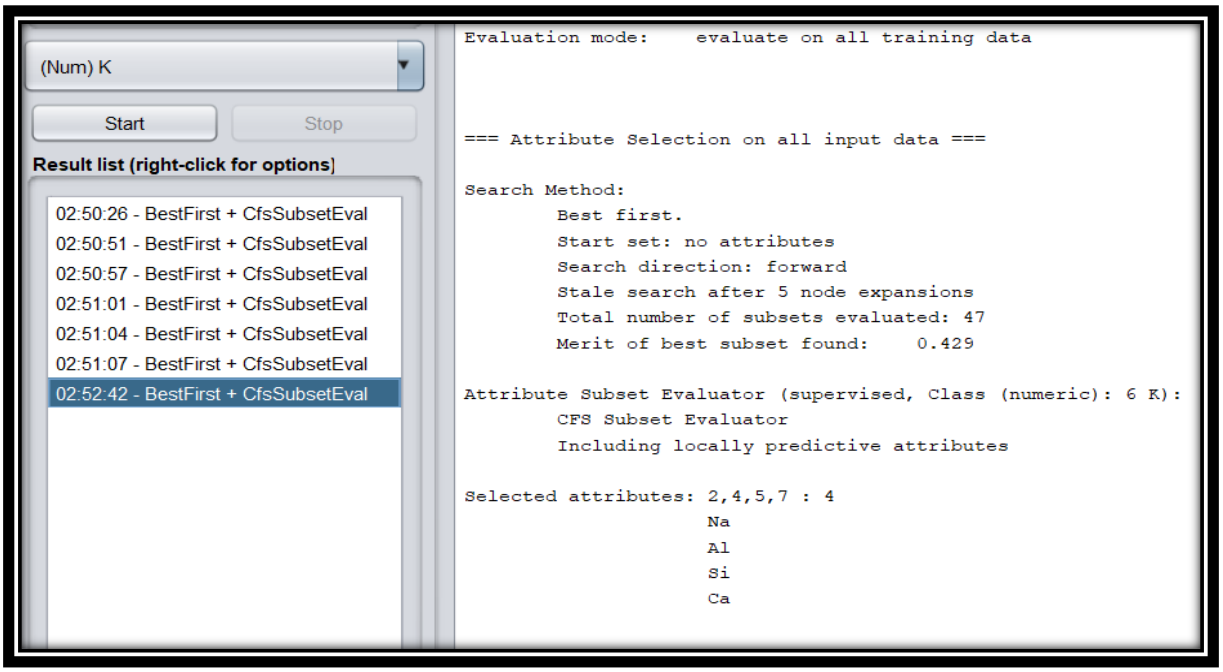
D. Genera un entrenamiento previo con los datos: Donde nos dice que tiene una predicción de 214 datos logrando clasificar en el primer cluster 88 que equivale al 44% del total del corpus, y el 126% que equivale al 56%.



Aquí se realiza la selección de atributos, donde nos indica que el uno , dos,  tres y el 6 son los datos que realmente tienen importancia, donde el 4 es el más importante,




Volvemos a seleccionar el segundo atributo, K y nos tendría que dar el resultado de 4, ya que nuestro interés ha sido analizar la importancia que tienen cada uno de los elementos en la producción del vidrio.



Se puede ver en la gráfica, los datos del corpus con la grafica, la grafica nos indica que elementos son los mas importantes. Gráficas 2D que relacionan pares de atributos.