

Tecnología automática para la clasificación y mejor destino de los crudos cubanos

Roberto Ayllón-Torres, Osvaldo Rodríguez -Morán, Carlos Morón-Alvarez

Centro de Investigaciones del Petróleo (CEINPET). Churrucá, No.481, e/ Vía Blanca y Washington, El Cerro.
La Habana. ayllon@ceinpet.cupet.cu, ormoran@ceinpet.cupet.cu.

Recibido: 6 de noviembre de 2014.

Aceptado: 6 de julio de 2015.

RESUMEN. La clasificación de los petróleos cubanos es una gran necesidad, y más aún, la determinación de su destino, para hacer más eficiente su utilización. Se propuso entonces como objetivo, llevar a cabo un procedimiento para obtener una clasificación de los crudos nativos. Para esto se analizó la naturaleza de los datos de los crudos nacionales y sus atributos, la cual constaba inicialmente de 30 atributos con 54 muestras de petróleo y se determinó procesarlos como atributos dividido por intervalos. Se aplicaron las herramientas del reconocimiento de patrones para clasificar los crudos en los cinco grupos establecidos: Ligeros, Medianos, Semi-Pesados, Pesados y Extra-Pesados. Se seleccionaron los atributos más representativos a partir del Peso Informativo. Se reconocieron cinco casos que estaban en contradicción con sus ubicaciones experticias iniciales en la Matriz de Aprendizaje, estas eliminaciones fueron justificadas. Se utilizó la similitud estricta por Intervalos, como medida de “cercanía” entre los casos. La magnitud de la Matriz de Aprendizaje resultó de 22 atributos y 49 casos. Finalmente, se diseñó un programa automático para la clasificación de los crudos y se obtuvo un 98.0% de eficacia en la clasificación de la Matriz de Control de 49 muestras de crudos nativos.

ABSTRACT. It is a very important the classification of the Cubans petroleum, and stiller, the determination of its destination, to make more efficient its use. The objective of this paper is to carry out a procedure to get a classification of the Cuban petroleum. For this it was analyzed the nature of the national data of the raw ones and their attributes, which consisted initially of 30 attributes with 54 samples of petroleum and it was determined to process them as attributes divided by intervals. The recognition patterns tools were applied to classify the Cuban petroleum in five established groups: Slight, Medium, Semi-heavy, Heavy and Extra-heavy. The most representative attributes were selected starting from the Informative Weight. Five cases were recognized that were in contradiction with their defined locations initially in the Learning Matrix, these eliminations were justified. The strict similarity was used as proximity measure among the cases. The magnitude of the Learning Matrix was of 22 attributes and 49 cases. Finally, an automatic program was designed for the classification of the raw ones and 98.0% of effectiveness was obtained in the classification of the Control Matrix of 49 samples of Cuban petroleum.

INTRODUCCIÓN

El petróleo es uno de los recursos naturales altamente valorado, desde todos los puntos de vista: económico, social, político y estratégico. De ahí que se incluya entre los minerales fundamentales en que se sostiene y depende la evolución del hombre actual.

La explotación, procesamiento y uso racional de dichos energéticos adquiere una connotación social de incalculable importancia (Ayllón *et al.*, 2000, Yumirka, *et al.* 2009).

En la actualidad se realizan esfuerzos con la finalidad de darle un uso racional a los petróleos nacionales ya sea para su procesamiento en las instalaciones de refinación o su combustión para la generación de energía. La industria petrolera necesita de un procedimiento automático de clasificación eficaz para los petróleos de Cuba junto con la propuesta de su uso racional y concretamente los crudos de la Franja Norte de Crudos Pesados (FNCP), entonces nos

preguntamos: ¿cómo instrumentar una tecnología automática para lograr una clasificación eficaz de los crudos nacionales cubanos además de proponer racionalmente su utilización?.

Como objetivo general está el desarrollar una tecnología automática para la clasificación y mejor destino de los crudos cubanos.

Existen varias definiciones de Modelos. En su sentido más amplio: un modelo abstrae desde cierto aspecto de la realidad. Identifica objetos, relaciones, propiedades y atributos de las entidades, relaciones entre propiedades de objetos (Steels, L., 1992). La inteligencia surge de la capacidad de construir y usar modelos, es decir de abstraer. Para Karbach *et al.* (1990), es una poderosa abstracción que permite reducir la complejidad de la realidad para focalizarse en ciertos aspectos.

Puede decirse que la modelación matemática “es el proceso de imitar la realidad utilizando el lenguaje de las matemáticas” (Karbach, W., Linster, M., Voss, A., 1990), este lenguaje es lo suficientemente rico como para tratar los más diversos problemas tanto por la materia de su contenido como por su complejidad. El procedimiento de construcción de un modelo matemático efectivo requiere de habilidad, creatividad y evaluación objetiva (Mendenhall, W.; 1994; Ripley, B. 1996; Therrien, C. W. 1989).

MATERIALES Y MÉTODOS

Existen varios criterios sobre las fases o etapas del proceso de modelación, por su simplificación expondremos la propuesta siguiente (Schulcloper, J. R. 1990; Schulcloper, J. R., *et al.*, 1999): Formulación del Problema, Formulación del Modelo, Selección del Modo de Solución, Solución del Modelo y Verificación del Modelo.

Debe tenerse en cuenta que un modelo no es la realidad, sino una representación de ella, es decir, con él estudiamos algunas de las características del fenómeno y la solución que nos dará tiene que ver con estas, si queremos estudiar otras entonces se debe reformular el problema y el modelo. Si el modelo es más refinado proporcionará más información pero la verdadera prueba es su capacidad de encontrar una respuesta aceptable para el problema planteado.

En el proceso de resolución de problemas en que interviene la modelación matemática pueden cometerse dos tipos de errores, uno es tratar de resolver una situación compleja por métodos empírico intuitivos ofreciendo una solución burda a un problema que requería un modelo; el otro es, por el contrario, tratar de aplicar un modelo a la solución de un problema para el cual no es requerido.

El petróleo es una mezcla compleja de hidrocarburos que contiene principalmente carbono e hidrógeno y compuestos con heteroátomos tales como: azufre, nitrógeno, oxígeno y bajas concentraciones de constituyentes metálicos.

La complejidad del petróleo y sus derivados se incrementa con el número de carbonos. La gasolina tiene un bajo número de componentes así como el combustible diesel un número yor.

Los constituyentes de los hidrocarburos pueden ser agrupados en hidrocarburos saturados, insaturados, y aromáticos.

Una de las desventajas que encuentran los expertos cuando clasifican los diferentes tipos de petróleos comparados, por ejemplo con la clasificación de otros combustibles como el carbón, estriba en que su composición elemental no se reporta con la misma extensión o amplitud. Las proporciones de los elementos en un petróleo varían, de forma ligera, en intervalo estrecho según el caso, con independencia de la amplia variación en las propiedades físico-químicas que abarcan desde los crudos más ligeros, de mayor movilidad a los bitúmenes asfálticos, en el otro extremo.

Los métodos de clasificación constituyen elementos indicativos de las características generales de un tipo de petróleo.

Uno de los sistemas que más se utiliza es el grado API. Por ejemplo, dentro de los crudos de un área específica, un crudo de 40 grados API (gravedad específica 0.825 g/cm^3) le corresponde, por lo general, un valor mayor que otro de 20 °API (densidad a $15^\circ\text{C} = 0.934 \text{ g/cm}^3$) porque contiene mayores cantidades de fracciones ligeras y menores de constituyentes pesados.

Con las consideraciones anteriores se pretendió reunir los aspectos que consideramos más importantes sobre el proceso de modelación. El problema se enmarca en la necesidad de establecer las diferencias, desde el punto de vista lógico-combinatorio, de los distintos tipos (o clases) de petróleos y dentro de ellos los de la Franja Norte de Crudos Pesados en términos de los atributos más importantes. La capacidad de clasificar un tipo de petróleo determinado en función de sus atributos es un aspecto de suma importancia en el proceso de exploración petrolera, industrias del petróleo y refinación, pues se podrá junto a otros elementos precisar los sistemas petroleros auténticamente naturales, además, desde el punto de vista de su refinación nos permite seleccionar que petróleos serán destinados para la producción de combustibles, obtención de aceites básicos lubricantes o diferentes tipos de asfalto de pavimentación.

Los datos originales de partida constan de 54 objetos y 22 atributos. Las categorías para la clasificación de los petróleos nacionales son las siguientes: Clase No. 1 (Petróleos ligeros), Clase No. 2 (Petróleos medianos), Clase No. 3 (Petróleos semi-pesados), Clase No. 4 (Petróleos pesados) y Clase No. 5 (Petróleos extrapesados).

Los atributos fueron obtenidos a partir de determinaciones de laboratorio, con lo cual podemos conocer el error de su determinación. Los atributos originales fueron: Azufre, Gravedad °API, Densidad(P15), $[\text{g/cm}^3]$, Densidad(P20), $[\text{g/cm}^3]$, Sulfuro de hidrógeno, $[\% \text{ m/m}]$, Agua por destilación, $[\% \text{ v/v}]$, Carbón Conradson, $[\% \text{ m/m}]$, Cenizas totales, $[\% \text{ m/m}]$, Asfaltenos, $[\% \text{ m/m}]$, Azufre total, $[\% \text{ mm}]$, Viscosidad cinemática, $[40^\circ\text{C}, \text{mm}^2/\text{s}]$, Viscosidad cinemática, $[100^\circ\text{C}, \text{mm}^2/\text{s}]$, Sedimento por extracción, $[\% \text{ m/m}]$, Nitrógeno, $[\% \text{ m/m}]$, Vanadio, $[\text{mg/kg}]$, Níquel, $[\text{mg/kg}]$,

Carbono, [% m/m], Hidrógeno, [% m/m], Factor de caracterización (Koup), Sales, [mg/l], Número total de ácido (TAN), [mgKOH/g], Punto de inflamación, [°C] y Punto de fluidez, [°C]

Formulación del Modelo.

Una partición del recorrido de los atributos se realiza, cuando la naturaleza de los datos muestra un mejor despliegue a partir de una agrupación por intervalos de clase.

Sobre la base de largas pruebas realizadas, se consideró conveniente procesar cada uno de los atributos (V_i) de forma intervalizada (k -valente), dentro de cada uno de los dominios correspondientes (\mathfrak{R}_i). Esto resultó así, porque el error de determinación de los atributos hizo conveniente trabajar con atributos intervalizados (Schulcloper, J. R., 1990). Significa que utilizaremos en vez del valor de un atributo determinado, el valor ordinal del intervalo al cual pertenece este atributo, esto es: se utilizaría K por V_i , si $V_i \in [I_k, I_{k+1}]$, donde I_1, I_2, \dots, I_m son los límites de los intervalos dentro del dominio \mathfrak{R}_i del atributo V_i .

La importancia de cada uno de los atributos se estimará a partir del llamado Peso Informativo (PI) (Schulcloper, J. R. 1990; Schulcloper, J. R., *et al.*, 1999).

Partimos de la siguiente pregunta: ¿cuál sería –para un atributo (V_i)– la magnitud del intervalo que dividiría su recorrido (\mathfrak{R}_i), de manera que separe mejor las clases involucradas? Esta determinación está dada precisamente, por la medida en que esta partición describa lo mejor posible los objetos involucrados en la Muestra de Aprendizaje (Tabla de Objeto-Propiedades.**TOP**), la cual por su magnitud no se muestra en este artículo.

La Entropía [1] ofrece la magnitud de la homogeneidad de la distribución de un atributo en su dominio y será el índice que nos apoyará para conocer cuál sería el intervalo que dividiría su recorrido (\mathfrak{R}_i), de manera que separe mejor las clases involucradas (Vladimirovich, *et al.*, 1986; Schulcloper, J. R. *et al.*, 1999).

$$E = - \sum_{i=1}^m p_i \log(p_i) \quad [1]$$

Donde:

M: número de grupos o intervalos.

p_i : probabilidad de ocurrencia para un intervalo i -ésimo.

Significa que si todos los valores se agrupan en un solo intervalo, entonces la Entropía tomará el valor extremo de cero. Por el contrario, si la distribución de las ocurrencias es igual para todos los intervalos, entonces esta tomará el valor máximo. En nuestro caso, todos los valores de Entropía serán normalizados respecto al valor máximo, por lo que los valores estarán en el intervalo [0,1]. De esta manera se podría establecer una simulación entre diferentes longitudes de intervalos para un atributo determinado y el valor de Entropía correspondiente para cada atributo. El intervalo para un atributo determinado se establecería, para el caso en que la Entropía sea mayor, es decir, que el atributo se distribuya lo más homogéneamente, por cada una de las clases consideradas.

Inicialmente tuvimos en cuenta dos expresiones para determinar la similitud entre dos objetos en función de sus atributos intervalizados, una a partir de la coincidencia estricta del intervalo correspondiente al atributo que se compara, la cual llamaremos “similitud estricta por intervalos” [2] y la otra a partir de la diferencia de los valores de los intervalos para un atributo, la que se designará como “similitud por diferencia de intervalos” [3].

Ambas expresiones, [2] y [3] tendrán valores en el intervalo entre cero y uno: cero (0), cuando no existe ninguna similitud y uno (1), cuando hay máxima similitud.

$$S(O_i : O_j) = \frac{\sum_{k=1}^{N_v} R_k}{N_v} \begin{cases} R_k = 0 & I_{i,k} \neq I_{j,k} \\ R_k = 1 & I_{i,k} = I_{j,k} \end{cases} \quad [2]$$

Donde:

$S(O_i : O_j)$ similitud entre los objetos.

i y j (=1 Similitud máxima; =0 Similitud mínima).

N_v es el número de atributos.

$I_{i,k}$ es el valor del intervalo del atributo.

k -ésimo del objeto.

i-ésimo.

$I_{j,k}$ es el valor del intervalo del atributo

o k-ésimo del objeto

j-ésimo.

R_k es el peso Informacional del atributo k-ésimo.

$$S(O_i : O_j) = 1 - \sum_{k=1}^{N_v} w_k \frac{|I_{i,k} - I_{j,k}|}{\Delta I_k - 1} \quad [3]$$

Donde:

$S(O_i : O_j)$

en la similitud entre los objetos.

i y j (=1 Similitud máxima; =0 Similitud mínima).

N_v es el número de atributos.

$I_{i,k}$ es el valor del intervalo del atributo.

k-ésimo del objeto.

i-ésimo.

$I_{j,k}$ es el valor del intervalo del atributo.

k-ésimo del objeto j-ésimo.

ΔI_k es la amplitud del intervalo de clase de la variable k-ésima.

W_k es el peso Informacional del atributo k-ésimo.

Para la confección del dendrograma, se utilizó el método de amalgamiento propuesto por Ward, por ser el más representativo a la naturaleza de nuestros datos y minimiza el número de casos “contaminantes”, es decir que clasifiquen en una clase a la que no pertenecen. Este método, minimiza la suma de cuadrados de cualquier dos grupos que pueden ser constituidos en cada paso de aglomeración (Ward, J. H. 1963). En general, este es un método muy eficiente.

Para definir el clasificador supervisado partiremos de las siguientes consideraciones: si O_i es un objeto ($i=1,2,3,\dots,n$), de atributos A_1, A_2, \dots, A_k y la clase C_o a la cual O_i pertenece ($C_o \in \{C_1, C_2, \dots, C_m\}$), entonces la clase a la que pertenece un objeto O_o que se quiere clasificar es C_o si:

$$S[O_i(A_1, A_2, \dots, A_k, C_m) : O_o] = \min \quad [4]$$

RESULTADOS Y DISCUSIÓN

En el proceso de selección de variables se aplicó el método de los pesos informativos (PI) (Schulcloper, J. R. 1990; Schulcloper, J. R., *et al.*, 1999). En la Tabla I, se muestran los pesos informativos para cada una de las variables, de estas, las que no aportan información son: Agua por destilación [%v/v], Carbón Conradson [%m/m], Asfaltenos [%m/m], Vanadio [mg/kg], Sales [mg/l], Número total de ácido (TAN) [mgKOH/g] y Punto de inflamación [°C].

Por ser los atributos de los crudos mediciones realizadas en el laboratorio, es que poseen errores en su determinación y esto llevó a que no se utilice el valor observado, sino de crear en el dominio del atributo una intervalización determinada y asignarle al valor observado, una asignación al intervalo correspondiente al cual pertenece y que finalmente, más adelante, facilite el proceso de clasificación.

Si para un atributo determinamos los valores de Entropía para cada valor de partición del recorrido de su dominio, tendremos entonces un arreglo de pares ordenados que en general se comportarían como se muestra en la Figura 1. Se puede apreciar, como por ejemplo en la Figura 1(a), que para el Nitrógeno, [%m/m], que la partición de 13 del recorrido, proporcionará una mejor representación de este atributo entre las clases. Para el caso de la Figura 1(b), Densidad P15, [g/cm³], la partición de 16 del recorrido, provocará una mejor representación de este atributo entre las clases propuestas. Es este valor de partición que tomamos para procesar todos los atributos como intervalizado (k-valente). Este proceso, para todos los atributos arrojaron los resultados mostrados en la Tabla II.

La utilización de los pesos informacionales de las variables se utilizará para disminuir la dimensión del espacio, eliminando los atributos que no contribuyen a la explicación del modelo, pero los testores asociados a ellos no lo utilizaremos en la clasificación por ser este trabajo una aproximación inicial al modelo definitivo, principalmente por la magnitud de los objetos, los cuales se pretende aumentar en trabajos futuros.

Tabla I. Relación de las variables y sus pesos informativos (PI) ordenados en orden creciente de estos.

Atributos	PI	Atributos	PI
Nitrógeno, [% m/m]	0.235	Carbono, [% m/m],	0.058
Cenizas Totales, [%m/m]	0.176	Hidrógeno, [% m/m],	0.058
Azufre Total, [%mm]	0.117	Factor de caracterización(Koup)	0.058
Sedimento por extracción, [% m/m]	0.117	Agua por destilación, [%v/v]	0.0
Gravedad [°API]	0.058	Carbón Conradson, [%m/m]	0.0
Densidad(P15), [g/cm ³]	0.058	Asfaltenos, [%m/m]	0.0
Densidad(P20), [g/cm ³]	0.058	Vanadio, [mg/kg],	0.0
Sulfuro de Hidrógeno, [%m/m]	0.058	Sales, [mg/l],	0.0
Viscosidad Cinemática, [40°C, mm ² /s]	0.058	Número Total de Acido(TAN), [mgKOH/g],	0.0
Viscosidad Cinemática, [100°C,mm ² /s]	0.058	Punto de Inflamación, [°C]	0.0
Níquel, [mg/kg]	0.058	Punto de Fluidez, [°C]	0.0

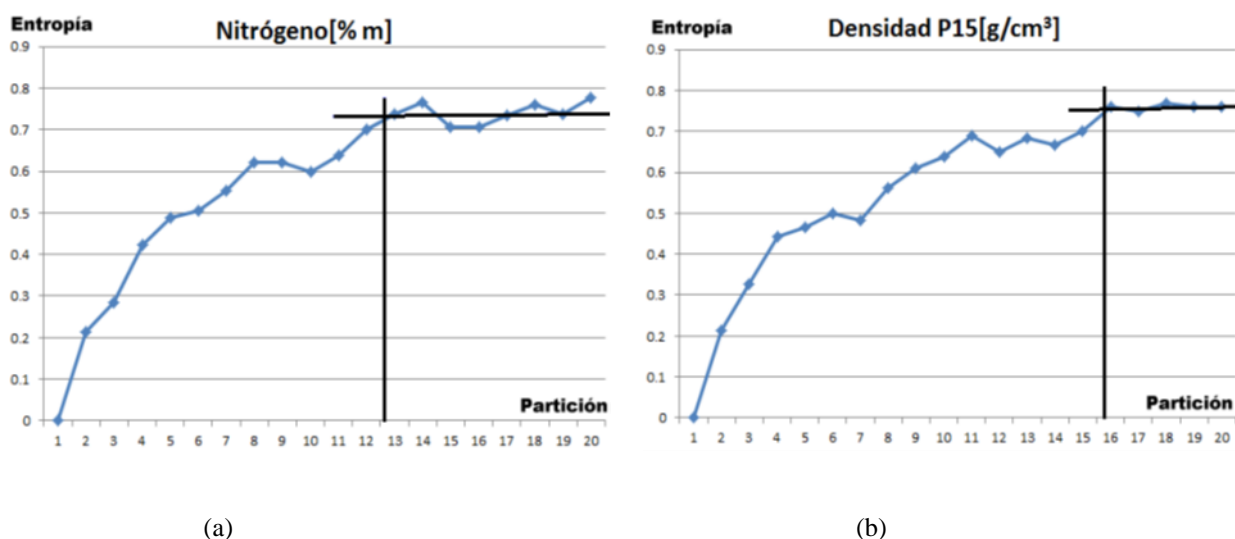


Fig 1. Ejemplo de gráficos de partición del recorrido de dos variables contra valor de la Entropía.

SOLUCIÓN DEL MODELO.

En la Figura 2 se aprecia el dendrograma de la Muestra de Aprendizaje (MA) para todos los atributos. Se puede ver en esta figura, que de forma muy eficiente, ya para un nivel normalizado de 20% respecto a la similitud máxima, los grupos que resultan están muy de acuerdo a lo propuesto inicialmente por el experto, es decir, en esta clasificación supervisada, los datos ofrecen inicialmente un alto grado de asociación entre los objetos de la misma clase. Las clases son: Clase No.1: Petróleos ligeros, Clase No.2: Petróleos medianos, Clase No.3: Petróleos semi-pesados, Clase No.4: Petróleos pesados y Clase No.5: Petróleos extra-pesados.

En el proceso de depuración de la Muestra de Aprendizaje, hubo que desestimar la participación de cinco objetos, para lograr una correcta la autoclasificación para la medida de similitud estricta por intervalos. La similitud diferencia de intervalos mostró un por ciento menor en la autoclasificación. Los objetos desestimados fueron:

- De la Clase No.5: Extra-pesados, el caso: Pozo Guanabo-100.
- De la Clase No.4: Pesados, los casos: Pozo Yumuri-220, Pozo Santa Cruz-302 y Punta Seboruco-16.
- De la Clase No.3: Semi-Pesados, el caso: Pozo Santa Cruz -102.

En el caso de los crudos de los pozos Santa Cruz-102 y 302, las muestras que se tomaron se encontraban en producción los mantos I y II de la Formación Veloz (que poseen tipos de petróleos diferentes), sin embargo, no se conoce cuál es el aporte de cada uno de dichos mantos. De ahí la no pertenencia de estos objetos en el grupo estimado de forma experticia.

Con respecto a los petróleos de los pozos Guanabo-100, Yumuri-220 y Seboruco-16, se señala que los mismos producen muy cercanos al contacto agua-petróleo por lo tanto los citados crudos se lavan con el agua de capa y parte de los componentes ligeros de ellos se arrastran por ella e influye en las propiedades físico-químicas de los mismos.

VERIFICACIÓN DEL MODELO.

Para la validación de la eficacia en la clasificación de los crudos de la Matriz de Aprendizaje, se utilizó una Matriz de Control de 49 muestras. Los atributos de estos crudos fueron analizados en el laboratorio y a partir de estos resultados recibió una clasificación del tipo de petróleo de cada muestra. Solo una de esas muestras, entregó una clasificación distinta a la esperada, por lo que podemos decir –en términos porcentuales- que la eficacia en la clasificación del modelo es de un 98.0%.

Tabla II. Resultado de aplicar el proceso de optimización del intervalo de clase para los recorridos de cada una de las variables.

No	Variable	Valor		Mejor Partición	Intervalo de clase
		Min	Max		
1	Nitrógeno [% m/m]	0.039	0.33	12	0.024
2	Cenizas Totales [% m/m]	0.003	0.15	14	0.011
3	Azufre Totales [% m/m]	0.55	9.92	13	0.721
4	Sedimento por extracción [% m/m]	0.0009	0.21	12	0.017
5	Gravedad [°API]	7.4	42.2	15	2.320
6	Densidad (P15) [g/cm ³]	0.8131	1.0446	15	0.015
7	Densidad (P20) [g/cm ³]	0.81	1.0414	13	0.018
8	Sulfuro de Hidrógeno [% m/m]	0	0.14	17	0.008
9	Viscosidad Cinemática [a 40°C,mm ² /s]	2.49	45504	14	3250.1
10	Viscosidad Cinemática [a 100°C,mm ² /s]	2	376.8	11	34.073
11	Níquel [mg/kg]	6	59	13	4.077
12	Carbono[% m/m]	79.77	93.05	15	0.885
13	Hidrógeno[% m/m]	10.01	14.77	13	0.366
14	Factor de caracterización (Koup)	9.5	11.6	14	0.150

Finalmente, se diseñó en la plataforma Microsoft Access, una base de datos, manipuladores y una aplicación para automatizar la clasificación de los petróleos a partir de la Matriz de Aprendizaje.

En la Figura 3 muestra la imagen del formulario para la captación de datos y la clasificación de los crudos nacionales. Cada uno de los elementos del formulario mostrado en la Figura 3, representan cajas de texto para mostrar el tipo de petróleo como resultado de la clasificación.

1. Caja de texto para colocar el nombre de la muestra de petróleo.
2. Cajas de texto de los atributos introducir.
3. Caja de texto para teclear las observaciones acerca del petróleo vigente.
4. Caja de texto que devuelve el tipo de petróleo (Resultado de la clasificación).
5. Cajas de texto para mostrar el objeto más parecido a la muestra clasificada.
6. Caja de texto cuyo valor indica el por ciento de influencia de los atributos en la clasificación.
7. Caja de texto en la se ha formado de forma textual los resultados de la clasificación.
8. Botón para generar el reporte breve de la muestra vigente.
9. Botón para generar el reporte ampliado de la muestra vigente.
10. Botón para generación de reporte diseñado por el usuario.
11. Botón para acceder (previa entrega de una clave) de los rangos de las variables.
12. Botón que adiciona un nuevo espacio para la introducción de una muestra nueva.
13. Botón que elimina de la base de datos la muestra vigente, previa confirmación de la acción.
14. Botones para el control y acceso de cada uno de los registros.
15. Caja de texto que muestra una ayuda en línea.
16. Botón para el abandono definitivo del sistema de base de datos.

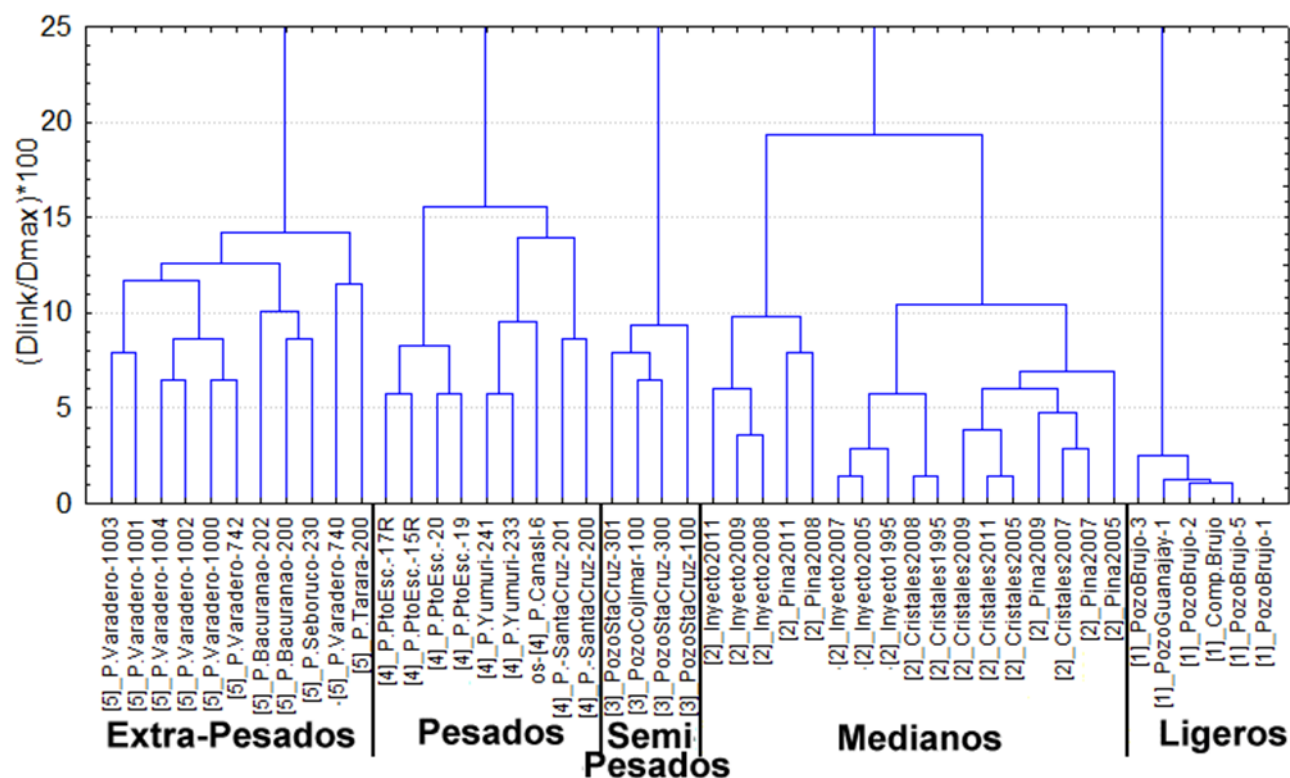


Fig 2. Dendrograma de la TOP intervalizada y la conformación de grupos para un nivel normalizado de 92% respecto a la similitud máxima.

En la Figura 4(a), se muestra la salida de un reporte breve de los resultados de la clasificación de un petróleo determinado y en la Figura 4(b), se muestra un reporte ampliado, el cual entrega al usuario los datos y resultados de una clasificación en la que se le informa, el posible destino del petróleo y su justificación.

The screenshot shows the 'Base de datos para la clasificación de petróleos cubanos' software interface. The interface is divided into several sections:

- Nombre de la muestra:** Compósito Jatibonico (1)
- Atributos:** A table of oil properties including Nitrogeno [% m], Gravedad Especifica [°API], Densidad_P15 [g/cm³], Sulfuro de Hidrógeno [% m], Cenizas Totales [% m], Azufre Total [% m], Densidad_P20 [g/cm³], Viscosidad Cinem [40°C, mm²/s], Viscosidad Cinem [100°C, mm²/s], Sedimentos por extracción [% m], Niquel [mg/Kg], Factor de caracterización (Koup), Carbono [% m], Hidrógeno [% m], and others (2).
- Observaciones:** A text box providing additional information about the oil sample (3).
- Clasificar:** A section where the user can select the oil type (Mediano) (4), the object (Pina2011) (5), and the influence of attributes (100%) (6).
- Conclusiones:** A section displaying the classification results and justification (7, 8, 9, 10, 11).
- Control de registros:** A section for managing the data records, including buttons for 'Nuevo registro' (12) and 'Elimina registro' (13), and a '6 de 49' indicator (14).
- Informe registro** (15), **Informe ampliado** (16), **Genera reporte**, and **Rango variables** buttons are also present.

Fig 3. Formulario para captación de datos, la clasificación de los petróleos y los respectivos reportes.

Reporte de la Base de Datos para la Clasificación de los Petróleos Cubanos.	
Nombre de la Muestra: Ligeros-[1] PozoBriego-1	
Rango TAPI 42	Resultados de la clasificación
Azufre 0.45	Tipo de Petróleo: Ligero
Grav. Específica [1]API 41.8	Objeto: PozoBriego-1
Densidad P25 [g/cm ³] 0.8131	% Influencia de atributos: 100 %
Densidad P20 [g/cm ³] 0.81	
Sulfuro de Hid. [w] 0.00005	Consideraciones
Agua e dest. [w] 0.05	La muestra Ligeros [1], PozoBriego-1, clasifica dentro del grupo de los petróleos: Ligeros.
Carbon Cond. [w] 0.07	Dentro de este grupo, se observó una cercanía al objeto PozoBriego-1. La influencia porcentual de los atributos, es de: 100%.
Cenizas T. [w] 0.05	
Arbitrenos [w] 0.16	
Azufre Tot. [w] 0.58	
Viscosidad Cinem [40°C, cm ² /s] 2.49	
Viscosidad Cinem [100°C, cm ² /s] 2.1	
Sed. & Extrac. [w] 0.001	
Hidrógeno [w] 0.06	
Vanadio [mg/kg] 19	
Níquel [mg/kg] 8	
Carbono [w] 83.45	
Hidrógeno [w] 12.46	
Factor Knop 11.55	
Salas [mg/l] 41.71	
TAN [mg/KOH/g] 0.01	
Pts. Inflam. -	
Pts. Fluidiz. -36	
Observaciones	

(a)

Reporte de la Base de Datos para la Clasificación de los Petróleos Cubanos.	
Nombre de la Muestra: Compuesto_Ambrosio	
Rango TAPI 14.5	Resultados de la clasificación
Azufre 0.93	Tipo de Petróleo: Mediano
Grav. Específica [1]API 41.8	Objeto: PozoBriego-1
Densidad P25 [g/cm ³] 0.8131	% Influencia de atributos: 100 %
Densidad P20 [g/cm ³] 0.81	
Sulfuro de Hid. [w] 0.00005	Consideraciones
Agua e dest. [w] 0.05	La muestra Compuesto_Ambrosio, clasifica dentro del grupo de los petróleos: Mediano. Dentro de este grupo, se observó una cercanía al objeto PozoBriego-1. La influencia porcentual de los atributos, es de: 100%.
Carbon Cond. [w] 0.07	
Cenizas T. [w] 0.05	
Arbitrenos [w] 0.16	
Azufre Tot. [w] 0.58	
Viscosidad Cinem [40°C, cm ² /s] 2.49	
Viscosidad Cinem [100°C, cm ² /s] 2.1	
Sed. & Extrac. [w] 0.001	
Hidrógeno [w] 0.06	
Vanadio [mg/kg] 19	
Níquel [mg/kg] 8	
Carbono [w] 83.45	
Hidrógeno [w] 12.46	
Factor Knop 11.55	
Salas [mg/l] 41.71	
TAN [mg/KOH/g] 0.01	
Pts. Inflam. -	
Pts. Fluidiz. -36	
Observaciones	

(b)

Fig 4. Ejemplo de salidas: Reporte breve (a) y Reporte ampliado (b).

CONCLUSIONES

Se pudo desarrollar una tecnología automática para la clasificación y mejor destino de los crudos cubanos se analizó la naturaleza de los datos de los crudos nacionales obteniéndose los pesos informativos correspondientes, los cuales permitieron disminuir la dimensión del sistema de 22 a 14 atributos y para cada uno de ellos, se obtuvo una intervalización del recorrido de su dominio que maximiza su representación discriminante frente a las clases establecidas. Se estableció un procedimiento de clasificación de los crudos nacionales utilizando dos medidas de similitud entre los crudos y concluyendo que la similitud estricta por intervalos ofrecía un mejor resultado en la clasificación. Se automatizó un procedimiento de clasificación de los crudos nacionales, el cual posee además un valor agregado de ser base de datos, lo que le imprime una memoria de las clasificaciones realizadas. Se evaluó la eficacia de la Muestra de Aprendizaje en la clasificación y mejor destino de los crudos nacionales utilizando una muestra de control de 49 objetos. La estimación de la eficacia en la clasificación, resultó de 98.0%.

RECOMENDACIONES.

Se recomienda seguir obteniendo nueva información, para continuar enriqueciendo la Matriz de Aprendizaje y disponer de más datos crear una Matriz de Control de un número mayor de crudos de la Franja Norte de Crudos Pesados, que apoyen el trabajo en la exploración y refinación de estos petróleos, de esta forma, y con un conocimiento mayor de la naturaleza de los datos, poder aplicar en la clasificación los testores típicos estimados.

REFERENCIAS BIBLIOGRAFICAS

- ASTMD-2422. (2012). Clasificación de lubricantes industriales por el sistema de viscosidad.
- Ayllón T. R.; Loginopoulo V.; Martínez Stable L. (2000). Proyecto de Servicio Especializado 200. Actualización de la calidad de los petróleos de las EPEP Occidente-Centro. Informe CUPET.
- Karbach, W.; Linster, M.; Voss, A. (1990). Model-Based approaches: One label-one idea? En Wielinga, Boose, Gaines Schreiber y Someren, Eds. *Current Trend in Knowledge Acquisition*. pp. 173-387. Amsterdam: IOS Press.
- Mendenhall, W.; Scheaffer, R.; Wackerly, D. (1994) Estadística Matemática con Aplicaciones. Grupo Editorial Iberoamericana.
- Ripley, B. (1996). Pattern Recognition and Neural Networks. Cambridge University.
- Schulcloper, J. R. (1990). Modelos Matemáticos para el Reconocimiento de Patrones. Facultad de Matemática, Física y Computación. Universidad Central de Las Villas. Folleto, p 245.
- Schulcloper, J. R.; Lazo, C. M.; Pico, P. R.; Sánchez, G. I. (1999). Workshop on Data Mining (Logical Combinatorial Pattern Recognition Group). Folleto, p 72.
- Steels, L. (1992). "Components of Expertise". *AI Magazine*, 11(2)pp. 28-49.
- Therrien, C. W. (1989) Decision, Estimation, and Classification. An Introduction to Pattern Recognition and Related Topics. Wiley, New York, NY.

10. Vladimirovich, L. P.; Ariosa, I. J. (1986). Búsqueda, Exploración y Evaluación Geólogo-Económica de Yacimientos Minerales Sólidos. Editora Pueblo y Educación. Tomo I, p. 348.
11. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, p 236.
12. Yumirka Comesaña, Y. G.; Dago-Morales A.; Talavera I.B.; Fernández R. F.; Diana Porro D.M. (2009). Comparación de dos métodos supervisados de reconocimiento de patrones para la clasificación de destilados medios de petróleo mediante Espectroscopia Infrarroja. Volumen 40. Número: 2. EDITORA.