



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE INGENIERÍA

Colegio de Ingeniería Geofísica

**Determinación de facies a partir de registros
geofísicos de pozo y propiedades petrofísicas
usando herramientas de inteligencia artificial**

TESIS

**PARA OBTENER EL TÍTULO DE
LICENCIATURA EN INGENIERÍA GEOFÍSICA**

PRESENTA:

Eduardo de Jesús Zurita Romero

DIRECTOR INTERNO:

M. C. Julio Cesar González Hernández

DIRECTOR EXTERNO:

Dr. Manuel Romero Salcedo

Puebla, Pue.

Enero 2020

En memoria de mi hermano

Abraham del Carmen

Zurita Romero

Todos los triunfos nacen cuando nos
atrevernos a comenzar

~ Eugene Wane

Agradecimientos

... A mis padres por toda su paciencia y apoyo. Por su constante empeño y esfuerzo en darme la mejor educación con el presente objetivo de que es la mejor inversión posible; y por apoyar mis locuras. Los amo más de lo que jamás podré expresar en mis años de vida. ...

A mi hermano Abraham y a mi hermana Lucia, que fueron mi apoyo; y que al mismo tiempo mis cómplices, confesores y delatores de todo lo que se puede vivir junto a alguien que no tuvo otra opción más que conocerte de toda la vida

A mi familia: tíos, tías, primos, primas, sobrinos, sobrinas y demás lazos sanguíneos en que no se está muy seguro de qué parentesco se tiene; sin ustedes, no sería lo que hoy soy, pues son parte de mi historia. ...

A mi asesor Dr. Manuel Romero; por aceptarme en las prácticas profesionales, que además de sus enseñanzas, lo considero un ejemplo a seguir y por embarcarse conmigo en esta monumental tarea.

A mi asesor Julio González Hernández; por haber elegido ser maestro, gracias maestro por haberme enseñado tan bien y por haberme permitido el desarrollo de esta tesis.

A la profesora y mi tutora Ileana Martínez, por trasmitirme todo el conocimiento para llegar hasta aquí y por preocuparse siempre por la generación 2014.

Al profesor Uribe, por transmitir su conocimiento en geología , por ser más que un profesor un amigo y aguantarme en todas sus clases.

Gracias a mi universidad, gracias por haberme permitido formarme y en ella, gracias a todas las personas que fueron participes de este proceso, ya sea de manera directa o indirecta, gracias a todos ustedes, fueron ustedes los responsables de realizar su pequeño aporte, que el día de hoy se vería reflejado en la culminación de mi paso por la universidad.

Este es un momento muy especial que espero, perdure en el tiempo, no solo en la mente de las personas a quienes agradecí, sino también a quienes invirtieron su tiempo para echarle una mirada a mi proyecto de tesis; a ellos asimismo les agradezco con todo mi ser.

..... Al Universo por permitirme llegar aquí, con todas las enseñanzas (buenas, malas, espantosas) que tengo y espero seguir acumulando. En serio, ¡vivir es tan divertido..!

Resumen

La industria petrolera mexicana requiere del desarrollo tecnológico para alcanzar sus metas productivas y enfrentar los retos de operar bajo entornos competitivos. En particular, bajo el contexto actual en el que una parte importante de la industria está enfocada en desarrollar actividades a lo largo de toda la cadena de valor de hidrocarburos que comprenden la exploración y extracción de los hidrocarburos líquidos y gaseosos.

Por décadas, la localización de reservas de hidrocarburos se ha logrado, gracias a la interpretación petrofísica entre otras técnicas, generando respuestas de los parámetros físicos que caracterizan a un estrato lo cual permite diferenciarlo de otros (comúnmente llamado facies). El objetivo de este trabajo de investigación es mostrar un método para obtener una clasificación de facies en rocas carbonatadas a partir de registros geofísicos convencionales y datos de microestructura de la roca. En esta investigación se pone en marcha una técnica para la clasificación de facies por medio de aprendizaje no supervisado, así como la construcción de modelos de predicción de facies a partir de una tecnología de inteligencia artificial, conocida como aprendizaje máquina (*Machine Learning*).

Los resultados obtenidos permiten determinar las facies presentes en el medio de estudio. Por lo tanto, con base a estos resultados se concluye que la metodología propuesta es la correcta y se proponen mejoras o modificaciones. Así mismo, se determinará la continuidad de los resultados de este trabajo en su acondicionamiento para utilizar en datos reales.

Abstract

The Mexican oil industry requires technological development to achieve its productive goals and face the operating challenges under competitive environments. In particular, under the current context in which an important part of the activities is focused on developing activities throughout the entire hydrocarbon value chain that include the exploration and extraction of liquid and gas hydrocarbons.

For decades, the exploration of hydrocarbon reserves has been achieved, thanks to petrophysical interpretation, generating responses of the physical parameters that characterize a stratum which allows differentiating it from others (commonly called facies). The objective of this research work is to show a method to obtain a classification of carbonate rock facies from conventional geophysical records and rock microstructure data. In this investigation, a technique for the classification of facies by means of unsupervised learning is launched, as well as the construction of prediction models of facies from an artificial intelligence technology, known as machine learning.

The results obtained allow to determine the facies present in the study medium. Therefore, based on these results, it is concluded that the proposed methodology is correct and improvements or modifications are proposed. Likewise, the continuity of the results of this work will be determined in its conditioning to be used in real data.

Contenido

Resumen	5
Abstract	6
Contenido	7
Lista de figuras.....	9
Lista de tablas.....	11
Capítulo 1. Introducción	12
1.1 Planteamiento del problema.....	13
1.2 Justificación.....	13
1.3 Objetivos.....	13
1.3.1. Objetivo general.....	13
1.3.2. Objetivos específicos.....	14
1.4 Hipótesis.....	14
Capítulo 2. Marco teórico	15
2.1 Teoría de facies.....	15
2.1.1 Tipos de facies.....	16
2.2 Rocas carbonatadas.....	19
2.2.1 Caliza.....	19
2.2.2 Dolomía.....	19
2.2.3 Origen de las rocas carbonatadas.....	20
2.2.4 Diagénesis.....	21
2.2.5 Dolomitización.....	21
2.2.6 Porosidad en las rocas carbonatadas.....	21
2.3 Petrofísica.....	25
2.3.1 Porosidad.....	25
2.3.2 Clasificación de la porosidad por su volumen.....	26
2.4 Registros geofísicos de pozos utilizados en este trabajo.....	27
2.4.1 Registro de densidad.....	28
2.4.2 Registro de neutrón.....	29
2.4.3 Registros rayos gamma.....	29
2.4.4 Registro de resistividad.....	32
2.4.5 Registros Sónicos de porosidad.....	33

2.5 Teoría detrás de la herramienta IMP-Predict V 2.0	34
2.5.1 Introducción	34
2.5.2 Aprendizaje de máquina y clasificadores.....	34
2.5.3 Máquina de soporte vectorial (SVM).....	35
2.5.4 Metodología del desarrollo del clasificador SVM	35
2.5.5 parámetros del clasificador SVM: C & γ	37
2.6 K-means Clustering en Python con Scikit-Learn	39
2.6.1 Descripción del algoritmo de K-means	40
2.6.2 Ventajas y desventajas del método K-means.....	41
2.7 Resumen del capítulo.....	42
Capítulo 3. Metodología	43
3.1 Introducción	43
3.2 Metodología del trabajo.....	43
3.3 Base de datos	45
3.4 Creación del código para la clasificación de facies	47
3.5 Descripción de facies generadas.....	51
3.6 Preparación de datos.....	55
3.7 Resumen del capítulo	56
Capítulo 4. Resultados: análisis de los modelos de predicción	57
4.1 Selección de los parámetros óptimos para el modelo de predicción.....	57
4.2 Modelos de predicción de la prueba 1: Datos Carbonatados.....	60
4.2.1 Descripción del intervalo oculto.....	60
4.2.2 Descripción de la predicción final dada en el intervalo 3	62
4.3 Modelos de predicción Prueba 2 : Facies generadas por asignación de porcentajes.....	64
4.3.1 Descripción de la predicción en el Intervalo oculto (Intervalo 1).....	64
4.3.3 Descripción de la predicción final dada en el intervalo 3	66
4.4 Comparación de resultados de la prueba 1 y 2	68
4.5 Modelos de predicción Prueba 3 : yacimiento carbonatado B.....	69
4.5.1 Descripción de la predicción dada en el intervalo 1 y 3.	69
Capítulo 5. Conclusiones y recomendaciones	73
Bibliografías	75
Anexos.....	78

Lista de figuras

Figura 1. Cambios laterales entre facies entre tres unidades de la misma edad (Tomada de Tarbuck & Lutgens, 2005).....	15
Figura 2. Ejemplo de una litofacies y biofacies (Tomada de Y.Yeples, 2015).	16
Figura 3. Fotografías de una microfacies de caliza con numulíticos y una microfacies de caliza con miliólidos.	17
Figura 4. Electrofacies generadas por el software IMP-Predict 2.0.....	18
Figura 5. Relación con los tipos de rocas carbonatadas.	20
Figura 6. Modelo de ambientes, sub ambientes y facies sedimentarias carbonatadas principales	22
Figura 7. Clasificación de porosidad en rocas carbonatadas	24
Figura 8. Clasificación general de la porosidad en porosidad primaria (parte inferior) y porosidad secundaria(parte superior)	26
Figura 9. Ejemplo de un registro combinado, graficado en IP SENERGY.....	31
Figura 10. Metodología para el desarrollo de un clasificador SVM.	37
Figura 11. Ejemplo gráfico de la clasificación realizada por el SVM de un conjunto de datos... ..	38
Figura 12. Clúster de datos con K=4.. ..	41
Figura 13. Esquema general del flujo de trabajo.....	44
Figura 14. Diagrama de flujo del método de clasificación de facies por medio del algoritmo de K-means.	48
Figura 15. Diagrama de flujo del método de clasificación de facies por medio de asignación de porcentajes.....	50
Figura 16. Graficas C y Gamma del modelo de predicción de los datos carbonatados	58
Figura 17. Resultados de la predicción del intervalo oculto (int1).	61

Figura 18. Resultados de la predicción del intervalo 3...	63
Figura 19. Resultados de la predicción del intervalo 1, prueba 2.....	65
Figura 20. Resultados de la predicción del intervalo 3, prueba 2.....	67
Figura 21. Resultados de la predicción del intervalo 1 yacimiento B.....	70
Figura 22. Resultados de la predicción del intervalo 2, yacimiento B - prueba 3.....	71
Figura 23. Acierto general de la predicción de la prueba 3.....	72
Figura 24. Pantalla de bienvenida de la herramienta computacional IMP-Predict Versión 2.0. .	79
Figura 25. Pantalla principal de la herramienta computacional IMP-Predict Versión 2.0.	79
Figura 26. Ventana de edición del proyecto predicción.....	80
Figura 27. Ventana de selección de parámetros.....	81
Figura 28. Ventana de selección de parámetros a predecir.....	81
Figura 29. Ventana de facies adyacentes.....	82
Figura 30. Ventana de personalización de colores para gráficos y nemónicos de facies.	83
Figura 31. Ventana de aviso de inicio del procesamiento.....	83
Figura 32. Ventanas de visualización de gráficas de C , Gamma y de ingreso de parámetros..	84
Figura 33. Ventana de aviso continuación del procesamiento.....	84
Figura 34. Despliegue de bitácora o registro de resultados.....	85
Figura 35. Ventana de confirmación de creación del reporte de resultados en PDF.....	86
Figura 36. Ventana de aviso de creación exitosa del reporte de resultados en formato PDF....	86
Figura 37. Gráficos C Y GAMMA generados de la prueba 1.....	91
Figura 38. Gráficos C Y GAMMA generados de la prueba 2.....	92
Figura 39. Gráficos C Y GAMMA generados de la prueba 3.....	93

Lista de tablas

Tabla 1. Densidad de los minerales más comunes (gr/cm ³).....	28
Tabla 2. Densidad de los fluidos más comunes (gr/cm ³).....	29
Tabla 3. Valores de velocidad e de compresión para materiales no porosos	33
Tabla 4. Información de las propiedades petrofísica del yacimiento de carbonatos A.....	45
Tabla 5. Información de los RGF de los datos del yacimiento de carbonatos B... ..	46
Tabla 6. Clasificación generada por medio de clúster (K-means) con las propiedades petrofísica del yacimiento de carbonatos A.....	52
Tabla 7. Clasificación de las facies reales del yacimiento de carbonatos B, con las variables de registro geofísicos de pozos.....	52
Tabla 8. Clasificación generada por medio de porcentajes de las propiedades petrofísicas.....	54
Tabla 9. Intervalos definidos en los datos de Excel para la pruebas 1 y 2 a procesar.....	55
Tabla 10. Intervalos definidos en los datos de Excel para la prueba 3 a procesar.....	55
Tabla 11. Pruebas para el procesamiento en IMP-Predict.....	56
Tabla 12. Plan de entrenamiento para el procesamiento en IMP-Predict con las 3 pruebas de los yacimientos de rocas carbonatadas.....	59
Tabla 13 Definición de facies, simbología y colores.....	59
Tabla 14. Acierto de las 12 pruebas generadas de la prueba 3.....	94

Capítulo 1. Introducción

Los registros de pozos son técnicas efectivas al estudiar las características de la estratificación del terreno (como apoyo a la exploración preliminar de nuevos campos). En la mayoría de las ocasiones en la industria del petróleo no se tiene toda la información completa de un pozo, es por ello que se deben utilizar diversas técnicas para entender el comportamiento de cierta litología en una profundidad dada. Debido a esto, es necesario el desarrollo tecnológico para alcanzar metas productivas y enfrentar los retos de operar bajo entornos complicados, en particular bajo el contexto actual en el que una parte importante de las actividades está enfocada en desarrollar la exploración y extracción de combustibles fósiles.

Los modelos físicos de rocas han evolucionado desde su implementación, hasta convertirse en parte integral de los proyectos de exploración e interpretación de los registros geofísicos de pozos. Actualmente, estos modelos se aplican para tratar de predecir las propiedades litológicas y petrofísicas de los yacimientos de hidrocarburos con el uso de inteligencia artificial. Con ello, la proliferación de estos modelos en los últimos años, han aumentado su confiabilidad como herramientas en las interpretaciones geológicas. Es por ello que surge esta propuesta de investigación en la que se plantea identificar sistemáticamente las facies basándose en las relaciones cuantitativas entre las propiedades físicas de las rocas y estudios geofísicos de pozo. Se mostrará el uso de la herramienta IMP-Predict, la cual permite construir modelos de predicción, basados en aprendizaje máquina (ML, por sus siglas en inglés), para determinar facies a partir de un conjunto de parámetros utilizados como variables predictoras (ej. registros geofísicos de pozos, propiedades petrofísicas), obteniendo como resultado las facies que no se tienen en una cierta profundidad del pozo.

1.1 Planteamiento del problema

En la industria del petróleo no se tiene toda la información descriptiva de un pozo, es por ello que se deben utilizar diversas técnicas para entender el comportamiento de cierta litología en una profundidad dada. El estudio y análisis de facies es fundamental para la predicción del comportamiento productivo del yacimiento, proceso orientado a la explotación óptima de la reserva de hidrocarburos.

1.2 Justificación

La realización de esta investigación permite dar el conocimiento y la efectividad de una herramienta de predicción de facies que brinda información geológica, partiendo de datos conocidos. Lo anterior da la pauta a diferentes comportamientos litológicos en una cierta profundidad.

De esta forma, ayudará al geofísico a dar la caracterización precisa de un yacimiento, no solo involucrando uso de herramientas geofísicas, sino también utilizando y analizando modelos de predicción para establecer las tendencias del comportamiento petrofísico, por medio del análisis integral de la información.

1.3 Objetivos

1.3.1. Objetivo general

Realizar la predicción de facies a partir de parámetros de microestructura de la roca y registros geofísicos convencionales para contribuir a una evaluación de la funcionalidad de la herramienta IMP-Predict y proporcionar información a los trabajos de ingeniería futuros.

1.3.2. Objetivos específicos

1. Implementar una técnica orientada a la clasificación de facies por medio de un algoritmo de aprendizaje no supervisado para obtener de una manera rápida y eficaz la clasificación.
2. Proponer una clasificación de facies con los datos petrofísicos para localizar el tipo de roca predominante con sus porosidades presentes en un yacimiento de rocas carbonatadas.
3. Reafirmar el potencial de la herramienta IMP-Predict en el análisis y predicción de facies, mostrando diversos procesamientos de datos, para así dejar la pauta de sus funciones en la industria.

1.4 Hipótesis

La carencia de información geológica y geofísica, han llevado a realizar este estudio para tener conocimiento de las técnicas de inteligencia artificial, por lo que:

- Tras la aplicación del código de clasificación y la herramienta IMP-Predict, se predecirá información sintética de pozos que servirá para validar la contribución geofísica en el mejoramiento en interpretación de los modelos de facies actuales, reduciendo la ambigüedad y posibles discrepancias entre las interpretaciones geológicas y geofísicas, proporcionando mayor certidumbre a la interpretación geológica del subsuelo.

Capítulo 2. Marco teórico

2.1 Teoría de facies

Cuando se estudia una serie de capas sedimentarias, se pueden ver los cambios sucesivos de las condiciones ambientales que hubo en un lugar concreto con el paso del tiempo. Esto es así debido a que, pueden existir muchos ambientes sedimentarios diferentes a lo largo de un terreno. Cada unidad tiene un conjunto de características que reflejan las condiciones de un ambiente en particular (Tarbuck & Lutgens, 2005).

El termino facies cubre un significado general, descriptivo y corresponde al aspecto que presenta una roca sedimentaria por una combinación particular de litologías, texturas, estructuras sedimentarias (físicas y biológicas), geometría, distribución de paleo corrientes, contenido fosilífero y color (Martínez, A. 2005). Este cuerpo de roca está delimitado por encima, por debajo y lateralmente, a partir de cambios en la combinación de sus propiedades, lo cual es el resultado de la suma de características litológicas, estructurales y orgánicas que pueden ser detectadas en el campo y que distinguen esta roca de las que la rodean(Figura 1). Dependiendo de las características o factores de las facies se pueden dividir en diferentes tipos.

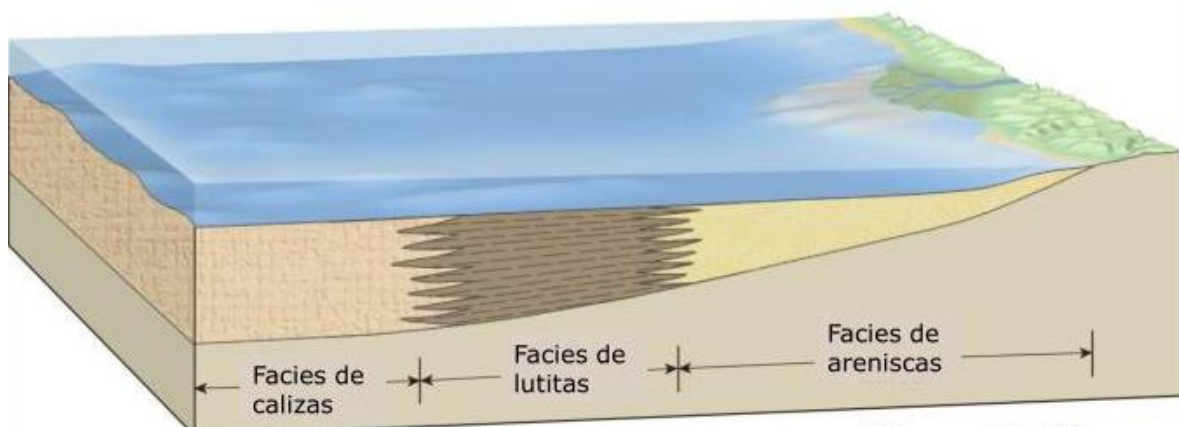


Figura 1.- La figura muestra los cambios laterales entre facies entre tres unidades de la misma edad (Tomada de Tarbuck & Lutgens, 2005).

2.1.1 Tipos de facies

Existen varios tipos de facies, debido a las diferentes características que se examinan en las rocas. De esta manera, términos como litofacies, biofacies y Microfacies, en la actualidad existen además las electrofacies (medidas en las digrafías) y las facies sísmicas (medidas en perfiles sísmicos). A continuación se describen los principales tipos de facies.

- Litofacies: se refiere a las características litológicas de un conjunto de estratos que son representativos de las condiciones físico-químicas que predominaron durante la depositación .
- Biofacies: se basan en las condiciones biológicas predominantes durante la depositación y a las características paleontológicas, por ejemplo facies de gastrópodos o facies de radiolarios. Es un término complementario de las litofacies (Figura 2).



Figura 2- Ejemplo de una litofacies y biofacies (Tomada de Y.Yeples, 2015)

- Microfacies: alude al conjunto de características paleontológicas, mineralógicas y petrofísicas examinadas en el microscopio, que se asocian a las condiciones genéticas que controlaron su depositación (figura 3).

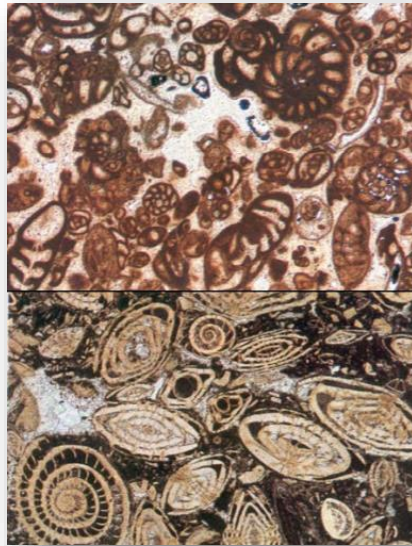


Figura 3.- La fotografía de arriba muestra una microfacie de caliza con numulíticos, y la fotografía de abajo nos muestra una microfacie de caliza con miliólidos (Tomada de A.E. Adams et al 1997).

- Electrofacies: se definen como el conjunto de respuestas de los parámetros físicos que caracteriza a un estrato y permite diferenciarlo de otros (Serra, 1982). El objetivo del análisis de electrofacies mediante los perfiles de pozo es describir objetivamente las formaciones atravesadas durante la perforación mediante sus respuestas en los perfiles y reconocer las distintas electrofacies fundamentales, así como analizar su asociación vertical con otras secuencias y por lo tanto poder deducir su evolución lateral, es decir reconstruir los modelos de electrofacies que ayudaran a definir los ambientes sedimentarios. En la siguiente figura se tiene un conjunto de electrofacies dadas por los registros de pozo (Figura 4).

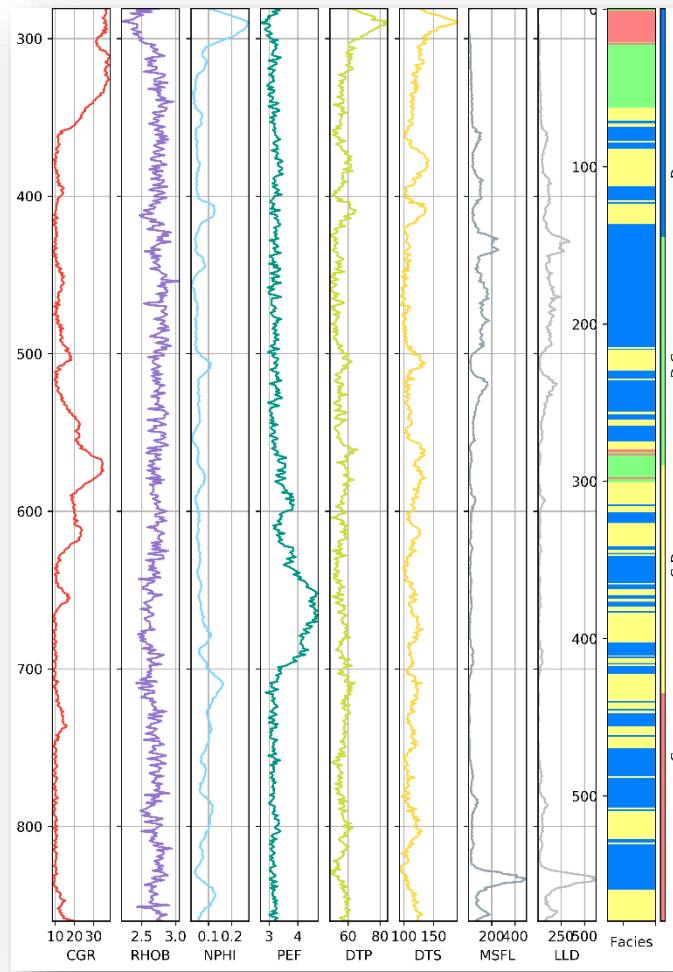


Figura 4. - Electrofacies generadas por el software IMP-Predict 2.0.

- Facies sísmicas: es la respuesta acústica de una litofacies que representa un sistema de depósito a una determinada señal acústica de características controladas. Por lo tanto está compuesta por reflexiones sísmicas, cuyos parámetros acústicos (amplitud, frecuencia, continuidad, intervalo de velocidad) difieren de los correspondientes a las unidades adyacentes dentro de una misma secuencia (Brown & Fisher, 1980).

2.2 Rocas carbonatadas

2.2.1 Caliza

La caliza es una roca compuesta por carbonato de calcio (CaCO_3). Son rocas originadas por un proceso de sedimentación que puede tener diversos orígenes, si bien la más importante la precipitación bioquímica, que se produce cuando el carbonato de calcio se fija en las conchas o esqueletos de determinados organismos, ya sean macroscópicos (braquiópodos, gasterópodos), microscópicos (foraminíferos) o nanoscópicos (cocolitos) y a la muerte de estos organismos, estas conchas o esqueletos se acumulan, originando un sedimento carbonatado. Cabe destacar que solo las calizas tienen un auténtico origen sedimentario, pues las dolomías se forman por procesos posteriores al depósito.

2.2.2 Dolomía

La dolomía ($\text{CaMg}(\text{CO}_3)_2$) es una roca sedimentaria de origen químico compuesta principalmente de dolomita. Para ser considerada como una dolomía esta debe contener 50% de dolomita como mínimo, ya que de acuerdo a la abundancia del mineral (calcita o dolomita) que haya en una roca se le asigna un nombre respectivo. Se considera una dolomía pura cuando existe más del 90% de dolomita. La dolomía es el producto de una diagénesis intensa, se forma por la precipitación directa de la dolomita pero también por la sedimentación de caparazones de organismos del plancton marino en los océanos. En la industria petrolera son de gran interés ya que por lo general resultan excelentes rocas almacenadoras de hidrocarburo. Cabe mencionar que las dolomías en comparación con las calizas, son rocas con mayor índice de fracturamiento (BIJU-DUVAL, 2002).



Figura 5.- Relación con los tipos de rocas carbonatadas (Tomada de Carbonate Reservoir Characterization, 1999).

2.2.3 Origen de las rocas carbonatadas

Los sedimentos carbonatados se producen comúnmente en aguas poco profundas o someras, en océanos cálidos, ya sea por precipitación directa del agua de mar o por extracción biológica de carbonato de calcio del agua de mar para formar material esquelético. El resultado es un compuesto de partículas de sedimento con una amplia gama de tamaños, formas, y mineralogías mezcladas entre sí para formar una variedad de texturas, composiciones químicas y, lo más importante, la distribución asociada al tamaño de poro (Lucia, 2007).

Se consideran rocas carbonatadas aquellas rocas que están formadas por más del 50% de minerales carbonatados. Las rocas carbonatadas se desarrollan a través de sedimentos biogénicos formados por actividad biológica, como la creación de arrecifes y la acumulación de restos de organismos en el fondo marino (Arche,2010).

2.2.4 Diagénesis

La diagénesis se define como el conjunto de procesos físicos, químicos y biológicos, que afectan los sedimentos después de su depositación, en el curso de su litificación durante el sepultamiento, bajo condiciones variables de tiempo, de temperatura ambiente y de presión, y por consiguiente produce cambios químicos y mineralógicos (BIJU-DUVAL, 2002). La importancia de estudiar y analizar la diagénesis, es por su interés económico, ya que durante este proceso ocurre una reducción en la porosidad y permeabilidad, lo cual influye directamente en la capacidad de almacenamiento de las rocas para contener ya sea agua, gases, petróleo o minerales. Las acumulaciones económicas están directamente relacionadas con indicios de buena porosidad y permeabilidad (Arche, 2010).

2.2.5 Dolomitización

Se define como el reemplazamiento de un sedimento o roca de carbonato de calcio por carbonato de magnesio. Se dan procesos de recristalización asociados, se crea porosidad secundaria, y hay destrucción de texturas de depósito. Las dolomías de reemplazamiento suelen ser más porosas y permeables que las calizas, lo que en parte explica su importancia como rocas almacenadoras de hidrocarburos.

2.2.6 Porosidad en las rocas carbonatadas

Se entiende como porosidad a la capacidad de las rocas para contener fluidos y es el resultado de la relación entre el volumen de espacios vacíos sobre el volumen total de la roca.

$$\phi = \text{Volumen Vacío} / \text{Volumen Total}$$

Los carbonatos son rocas sedimentarias depositadas en ambientes marinos de aguas claras, someras o cálidas y en su mayoría poseen un origen biológico (Figura 6).

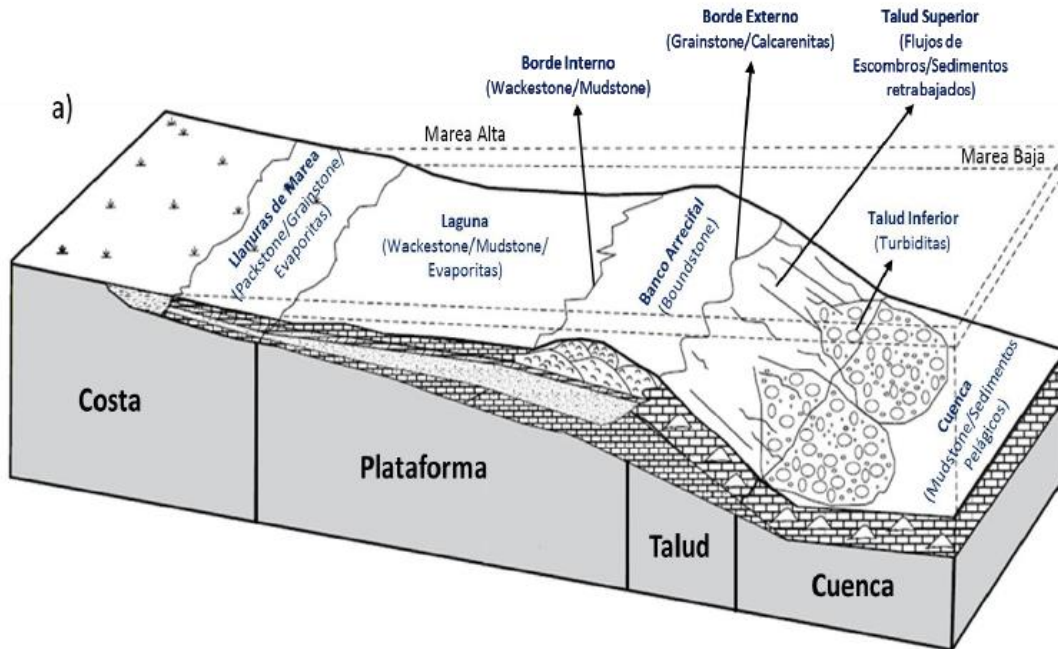


Figura 6.- Modelo de ambientes, sub ambientes y facies sedimentarias carbonatadas principales. (tomada de Hamid, et al., 2012).

Una diferencia clave entre las rocas clásticas y las rocas carbonatadas es la distancia existente entre el sitio donde se creó el sedimento y aquel en donde se depositó. Mientras que la sal o el limo pueden viajar cientos de kilómetros por los sistemas fluviales antes de su depositación y su litificación, los granos que componen los sedimentos carbonatados se depositan normalmente muy cerca del lugar en que se originaron. Esta depositación local contribuye significativamente a la heterogeneidad de los granos de carbonato. Una vez formada la roca carbonatada, una diversidad de procesos químicos y físicos comienza alterar la estructura de la roca modificando las características fundamentales, tales como la porosidad y la permeabilidad. En el momento de la depositación, los sedimentos carbonatados a menudo poseen porosidades muy altas (35%–75%); sin embargo, estos valores decrecen bruscamente a medida que el sedimento es alterado y sepultado en las profundidades del yacimiento. En consecuencia, los yacimientos carbonatados exhiben variaciones grandes y abruptas en la distribución de los tipos de roca (Tarbuck et al., 2005).

La porosidad de las rocas carbonatadas se agrupan en tres tipos:

1. Porosidad conectada: existente entre los granos de carbonato.
2. Vesículas: que son poros desconectados resultantes de la disolución de la calcita en el agua durante la diagénesis.
3. Porosidad por fracturamiento: causada por los esfuerzos presentes luego de la depositación.

Otra forma de agruparse desde el punto de vista petrofísico se puede dividir en: porosidad de matriz, porosidad de fractura y porosidad vugular. En muchos casos, los yacimientos carbonatados naturalmente fracturados poseen tres sistemas de porosidad: porosidad de fractura, porosidad de matriz y porosidad vugular (tanto conectada como no conectada), lo que implica la complejidad del modelado del yacimiento (Bratton, 2006).

- Porosidad de matriz: es igual al espacio vacío en la matriz respecto a un volumen total de roca. La porosidad de matriz incluye los sistemas de porosidad intercrystalina e intergranular o interparticular (Aguilera, 1995; Doveton, 2014).
- Porosidad vugular: se define como el espacio de poro que está dentro de los granos o cristales o que es significativamente más grande que los granos o cristales y puede ser conectada y no conectada. Los vúgulos están comúnmente presentes como granos disueltos, cámaras fósiles, fracturas y grandes cavidades irregulares. En la industria petrolera el término de porosidad vugular se utiliza para referirse al espacio de los poros visibles en las rocas carbonatadas.
- Porosidad de fractura: se define como las aberturas comúnmente creadas por fallas estructurales de las rocas del yacimiento, causadas bajo compresión o tensión por las actividades tectónicas como pliegues y fallas. Estas aberturas incluyen juntas, fisuras y fracturas. La porosidad debido a las fracturas en los carbonatos usualmente no supera el 1%. También se pueden crear microfracturas causadas por la presión confinante (Lucia, 2007).

CONDICIONADA POR LA FÁBRICA		NO CONDICIONADA POR LA FÁBRICA					
PRIMARIA		Interpartícula		Fractura			
		Intrapartícula			Canales		
		Fenestral				Huecos (Vug)	
		Shelter					Caverna*
		De crecimiento			*Caverna se aplica a los huecos o canales de tamaño igual o superior al de un humano.		
SECUNDARIA		Intercristalina					
		Mólida					
CONDICIONADA O NO CONDICIONADA POR LA FÁBRICA							
				De brechificación	Perforación	Galería	Grieta de retracción

Figura 7.- Clasificación de porosidad en rocas carbonatadas (Choquette y Pray,1970)
(tomada de Bachle & Harmersbach, 2009).

En resumen, las modificaciones texturales con frecuentes intercambios entre los espacios vacíos y la materia sólida es la norma en las rocas carbonatadas. En consecuencia, una determinada roca puede presentar varios tipos de poros, correspondientes a distintos procesos, a distintas generaciones o ambos hechos. En conjunto, estas tres formas de porosidad crean un camino muy complejo para los fluidos y afectan directamente la productividad de los pozos. Esta heterogeneidad produce además un impacto sobre la respuesta de las mediciones de los registros y por consiguiente, sobre la determinación del petróleo en sitio. Además de las variaciones producidas en la porosidad, la mojabilidad es una característica heterogénea adicional de los carbonatos. La caracterización de la distribución de la mojabilidad y la comprensión de sus efectos sobre el flujo de fluidos dentro de un yacimiento complejo es crucial para estimar las reservas recuperables y determinar las estrategias de producción destinadas a maximizar la recuperación.

2.3 Petrofísica

La petrofísica es una especialidad que involucra conocimientos de ingeniería del petróleo, geofísica y geología, la cual determina cualitativa y cuantitativamente las propiedades físicas y texturales de la roca, especialmente la distribución de los poros, que sirven como depósitos para las acumulaciones de hidrocarburo. Adicionalmente, determina la relación existente entre los fluidos y su movimiento a través del medio poroso de la roca de un yacimiento determinado, mediante la integración del entorno geológico, registros geofísicos de pozo, análisis de muestras de roca y sus fluidos e historias de producción (Lucia, 2007).

2.3.1 Porosidad

La porosidad de un yacimiento es definida como la fracción del volumen total del yacimiento que no es ocupada por la estructura sólida del yacimiento (Tiab & Donaldson, 2004). Esta puede ser expresada de forma matemática como:

$$\phi = \frac{V_b - V_{gr}}{V_b} = \frac{V_p}{V_b}$$

Donde:

ϕ = porosidad

V_b = volumen total de la roca de yacimiento

V_{gr} = volumen de grano

V_p = volumen de poro

La magnitud de la porosidad depende de los siguientes factores: uniformidad en el tamaño de grano, forma, textura de superficie, angularidad, orientación, grado de cementación y grado de compactación durante y después del depósito.

La porosidad puede dividirse desde tres puntos de vista: por su génesis o depósito (porosidad primaria y porosidad secundaria), por su volumen (porosidad total y porosidad efectiva o porosidad no efectiva) y por la conducción eléctrica o flujo de fluidos (porosidad de flujo y porosidad de entrapamiento).

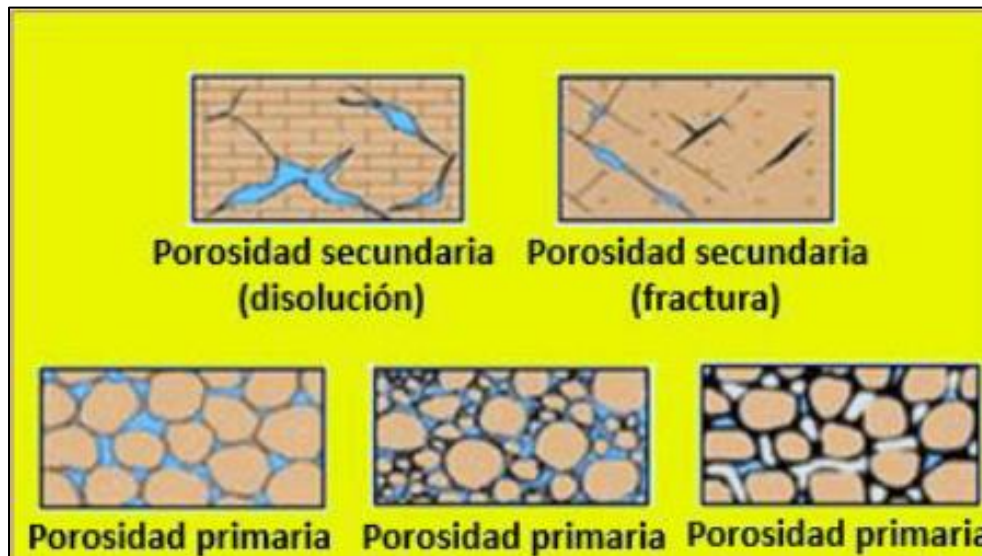


Figura 8.- Clasificación general de la porosidad en porosidad primaria (parte inferior) y porosidad secundaria (parte superior)(tomada de Tiab & Donaldson, 2004).

2.3.2 Clasificación de la porosidad por su volumen

- Porosidad total ($\emptyset T$): se define como el volumen poroso total que existe en la roca, esto es, los poros interconectados (matriz, vúgulos y fracturas) más la suma de los poros no conectados.
- Porosidad efectiva ($\emptyset e$): es aquella a la cual se le ha quitado el efecto de la arcilla e involucra el volumen de poros interconectados, y se puede representar matemáticamente de la siguiente manera:

$$\emptyset e = \emptyset T * (1 - Vsh)$$

$\emptyset e$ = porosidad efectiva

$\emptyset T$ = porosidad total

V_{sh} = volumen de arcilla

- Porosidad no efectiva (ϕ_{ne}): es aquella en la cual los poros no están interconectados entre sí o la conectividad es tan baja que impide que los fluidos se conduzcan de manera adecuada en las rocas.

2.4 Registros geofísicos de pozos utilizados en este trabajo

Los registros geofísicos de pozos son técnicas para determinar las diferentes propiedades de las rocas y consisten en la medición, en función de la profundidad o del tiempo de alguna propiedad física y/o química asociada al material litológico del subsuelo (sistema roca-fluido). El registro se lleva a cabo mediante una herramienta denominada sonda, la cual se compone de uno o más sensores que se encargan de medir sobre la pared del pozo la respuesta física del material litológico. Dicha respuesta medida está en función del tiempo. El diagnóstico que proporcionan es puntual, continuo de alta resolución, haciendo de los registros una técnica ampliamente utilizada en la industria petrolera, minera, geohidrología, geotérmica y geotécnica, para determinar saturación de fluidos. (Geotem, 2016)

En este trabajo se utilizaron los registros geofísicos convencionales:

- Registros acústicos – tiempos de onda P (DTP) y onda S (DTS)
- Registros de resistividad – micro-herramienta (MSFL) y resistividad profunda
- Registros de neutrones – porosidad total (NPHI)
- Registros de densidad – densidad de formación (RHOB)
- Registros de rayos gama (GR)

Donde por medio de un proceso de inversión con un software propio del instituto Mexicano del petróleo se obtuvo la información del tipo de roca con sus porosidades del yacimiento de rocas carbonatadas.

2.4.1 Registro de densidad

Los registros de densidad se utilizan para detección de gas, determinación de la densidad de hidrocarburos, evaluación de litologías complejas y para cálculos de propiedades mecánicas de las rocas. Este registro es de los más útiles en la industria petrolera ya que presentan un valor muy exacto de la porosidad efectiva en arenas arcillosas. En la tabla 1 se muestran datos de las densidades de los minerales más comunes de las rocas asociadas a los yacimientos de petróleo, en la tabla 2 los datos de los fluidos contenidos en este mismo tipo de yacimientos.

Mineral	Densidad ($\frac{\text{gr}}{\text{cm}^3}$)
Lignita	1.15-1.30
Carbón bituminoso	1.14-1.70
Antracita	1.30-1.70
Montmorillonita	2.0
Azufre	2.05
Halita	2.17
Yeso	2.32
Feldespatos	2.55-2.80
Cuarzo	2.65
Calcita	2.71
Dolomita	2.87
Anhidrita	2.95
Magnesita	3.0-3.12

Tabla 1. Densidad de los minerales más comunes (gr/cm³). (Tomada de Gómez R. M. & Ramírez H. R., 2004).

Fluido	Densidad ($\frac{gr}{cm^3}$)
Gas natural	0.00078
Aire	0.00129
Gas Natural a 210 Kg/cm ² y 32°C	0.146
Aceite (50° API)	0.780
Aceite (30° API)	0.876
Agua dulce	1.000
Agua de mar	1.026
Agua salada	1.15

Tabla 2. Densidad de los fluidos más comunes (gr/cm³). (Tomada de Gómez R. M. & Ramírez H. R., 2004).

2.4.2 Registro de neutrón

Los registros de neutrones, son utilizados en la delineación de formaciones porosas y así determinar las diferentes porosidades. Son respuestas de la cantidad de hidrógeno presente en formaciones. En formaciones limpias, donde los poros están saturados de agua o aceite, el registro responde con la cantidad de porosidad saturada en el fluido. Las zonas de gas con frecuencia se identifican al comparar el registro de neutrones con uno o más registros de porosidad e identificación de litologías aún más exactas (Schlumberger, 2008).

2.4.3 Registros rayos gamma

Como es bien conocido las formaciones geológicas emiten una cierta radiación, donde la magnitud de esta radioactividad depende de las características de la roca.

Es un registro de radioactividad de las rocas, producto de la desintegración natural de las pequeñas cantidades de elementos radioactivos que contienen, los elementos radiactivos tienden a concentrarse en arcillas y lutitas, por lo tanto la respuesta de la curva de rayos gamma naturales nos indica la diferencia entre uno y otro tipo de roca contenido en la formación (Figura 9).

Se concluye que la curva de rayos gamma naturales es un registro complementario en la interpretación de la litología de las formaciones atravesadas por un pozo, teniendo como objetivo el apoyar la interpretación del registro de potencial espontáneo, que de igual manera permite cualitativamente identificar zonas permeables, por lo que ambos se consideran registros correlacionables, debido a que las curvas generadas son similares. Los valores relativos de radiactividad según el tipo de roca se pueden clasificar de la siguiente manera:

1. Rocas de muy baja radiactividad: como el carbón, sal y anhidritas.
2. Rocas de baja radiactividad: entre estas se puede citar las calizas puras, dolomías, areniscas y arenas o mezclas de rocas (las arenas y areniscas de grano fino tienen mayor radiactividad que las de grano grueso).
3. Rocas de radiactividad media: se asocia a las areniscas y a las arenas arcillosa y limosas, calizas, dolomías arcillosas, lutitas arenosas y calcáreas.
4. Rocas de alta radiactividad: dentro de este grupo se tiene la arcilla, lutitas, cenizas volcánicas y bentonitas

De acuerdo con esta clasificación de rangos de radiactividad natural, las deflexiones de la curva de rayos gamma hacia la derecha indicara el aumento en la radiactividad. Como la lutitas es la roca más común casi siempre se podrá trazar una línea de base de lutitas que sirva de referencia, como si se tratara de una curva de potencial espontaneo a partir de esta línea se podrá identificar, las capas de menor y mayor radiactividad (Schlumberger, 2008).

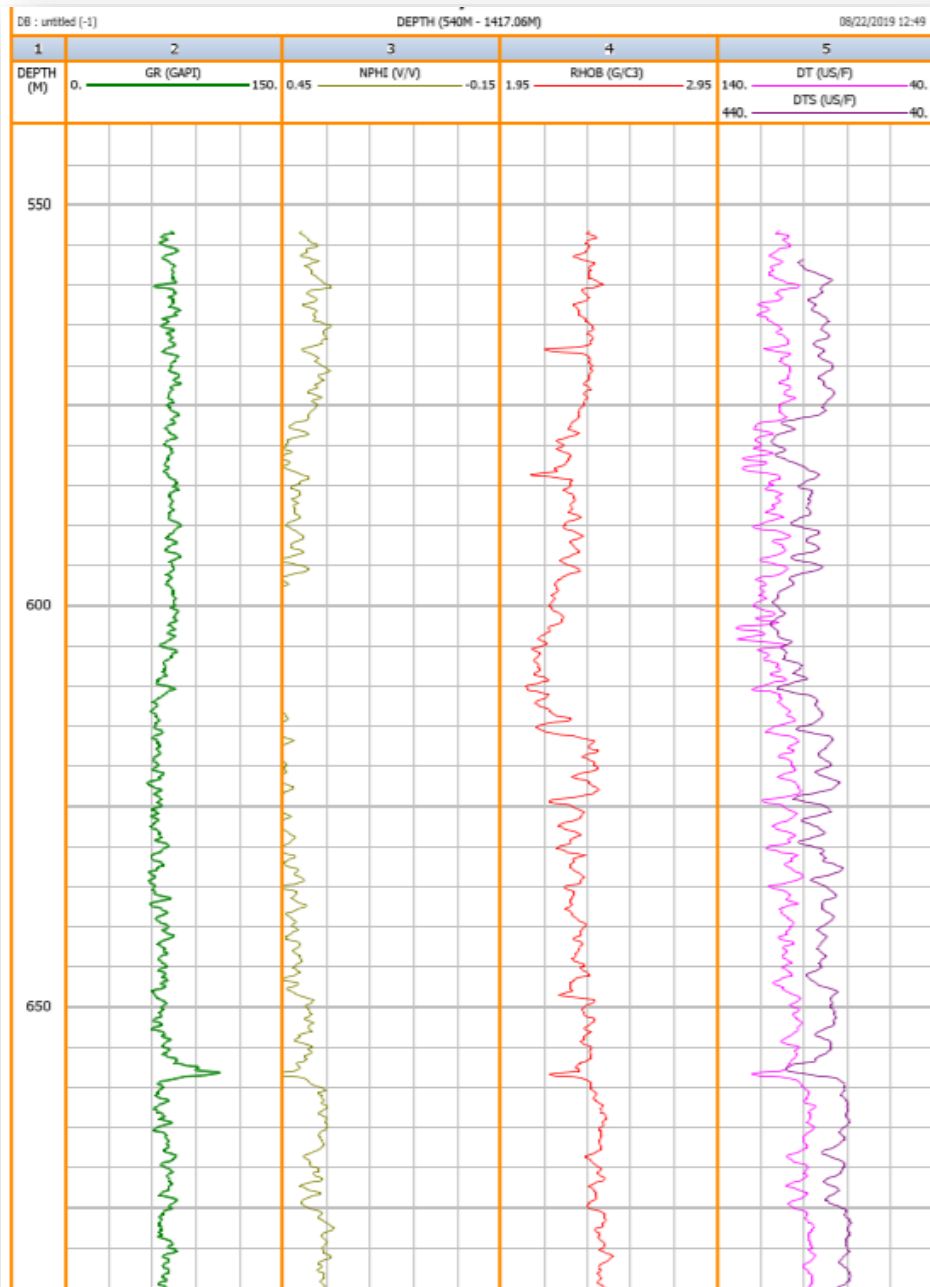


Figura 9.- Ejemplo de un registro combinado, donde el primer carril representa la profundidad, el segundo carril representa los registros de rayos gamma, el tercer carril representa los registros de neutrón porosidad, el cuarto carril representa la densidad total y el quinto carril los registros mecánicos, generados por software IP, SENERGY

2.4.4 Registro de resistividad

La resistividad es la capacidad que tienen las rocas de oponerse al paso de corriente, eléctrica inducida y es el inverso de la conductividad, se considera como la resistencia de un material conductor de 1 metro de longitud con área transversal de 1 metro cuadrado. La unidad de la resistividad es el Ohm-Metro ($\Omega\text{-m}$).

Para llevar la medición de esta propiedad en un pozo, se realiza un arreglo de electrodos que se introducen en el pozo, actuando el lodo como un medio conductor que liga los electrodos con las formaciones. Si los poros de una formación contienen agua salada presentará alta conductividad con resistividades bajas, pero si están llenos de petróleo o gas presentará baja conductividad y por lo tanto la resistividad será alta. Las rocas compactas poco porosas como las calizas masivas poseen resistividades altas.

Existen dos tipos principales de perfiles resistivos: el Perfil Lateral (*Laterolog*) y el Perfil de Inducción (*Induction Log*). El perfil lateral se utiliza en lodos conductivos (lodo salado) y el perfil de inducción se utiliza en lodos resistivos (lodo fresco o base aceite).

los Perfiles Laterales que existen son :

- MSFL= Microspheric Laterolog. Para las proximidades (1.0" y 6.0"). Lee la resistividad de la zona lavada (Rxo)
- MLL = LLM = Micro Laterolog. Para las proximidades (1.0" y 6.0")
- SLL = LLS = Someric Laterolog. Para profundidades someras (0.5" y 1.5")
- DLL = LLD = Deep Laterolog. Para profundidades de más de 3.0".

Miden resistividad de la formación (R_t). Se lee de izquierda a derecha, en escala logarítmica. La unidad de medida es el ohm-m, con un rango de valores que va desde 0.2 hasta 2000 $\Omega\text{-m}$ (Schlumberger, 2008).

2.4.5 Registros Sónicos de porosidad

Es un registro de la profundidad contra (ΔT), que es el tiempo requerido por una onda compresional de sonido para recorrer un pie de formación, conocido como tiempo de tránsito, el cual es el valor recíproco de la velocidad de la onda compresional del sonido.

El tiempo de tránsito en una formación dependerá de la litología y porosidad, a su vez el tiempo de tránsito sísmico es útil para realizar una interpretación de un registro sísmico.

Tratándose de formaciones de rocas sedimentarias, la velocidad del sonido se verá afectada por diversos factores, uno de ellos es el material de la roca de matriz, ya sea una arenisca, caliza o dolomita, otro factor a destacar es la porosidad distribuida, en la tabla 3 se presentan los rangos de valores de velocidad sísmica y del tiempo de tránsito (ΔT), para matrices de roca y revestimientos comunes.

Material	Velocidad sísmica (ft/sec)	Tiempo de tránsito ΔT ($\mu seg/ft$)
Aceite	4300	232
Agua(Lodo)	5000 – 5300	200 – 189
Neopreno	5300	189
Pizarras	6000-16000	167-62.5
Roca salada	15000	66.7
Arenisca	15000- 20000	55.6
Anhidrita	20000	50.0
Carbonatos	21000-23000	47.6-43.5
Dolomitas	24000	42

Tabla 3. Valores de velocidad e de compresión para materiales no porosos(tomada de Gómez R. M. & Ramírez H. R., 2004).

Los valores mencionados en la tabla anterior, son para sustancias no porosas, la porosidad disminuye a la velocidad del sonido a través del material de la roca y al mismo tiempo aumenta el tiempo de tránsito (Schlumberger, 2008).

2.5 Teoría detrás de la herramienta IMP-Predict V 2.0

2.5.1 Introducción

IMP-Predict es una herramienta computacional que permite construir modelos de predicción, basados en aprendizaje máquina (*Machine Learning*). Para determinar facies a partir de un conjunto de parámetros, utilizados como variables predictoras (ej. registros geofísicos de pozos, propiedades petrofísicas, atributos sísmicos). Se fundamenta en una técnica que comprende la exploración de los datos, su acondicionamiento, la creación del modelo de predicción, así como la evaluación y aplicación de dicho modelo para realizar la clasificación de facies.

2.5.2 Aprendizaje de máquina y clasificadores

El aprendizaje máquina o machine learning, constituye un área científica, posicionada entre las matemáticas, la estadística y las ciencias de la computación, de donde han emergido algoritmos capaces de aprender, extrayendo la estructura compleja de los datos y realizando tareas de predicción de propiedades, características o tendencias de datos desconocidos con fines de clasificación, agrupación (*clustering*) y regresión (Hall, B. 2016). El modelo de predicción que se construye está soportado por un método de aprendizaje supervisado, usado para la clasificación, llamado máquina de soporte vectorial (del inglés Support vector machine o SVM), mediante el cual se realiza la determinación de facies a partir de un conjunto de variables predictoras que pueden ser: registros geofísicos de pozo, propiedades petrofísicas o incluso atributos sísmicos. Al conjunto de variables predictoras se conoce como vector característico y al modelo de predicción se le conoce como clasificador SVM. Una vez obtenido el clasificador SVM se determinan las facies para un vector característico del conjunto de datos que no se le ha dado a conocer al clasificador, durante la etapa de entrenamiento, permitiendo así validar

el modelo y quedando listo para realizar la predicción de cualquier número de pozos cuyas facies sean desconocidas, dichos valores que se determinen con un clasificador SVM se les conoce como clases (Hall, B. 2016).

2.5.3 Máquina de soporte vectorial (SVM)

Un clasificador SVM es un tipo de algoritmo de aprendizaje supervisado, el cual requiere que sea alimentado con datos para llevar a cabo un proceso de entrenamiento con el fin de aprender de la estructura y relaciones que existen entre los datos de entrada (vectores característicos) y los valores de salida que se determinarán (clases). La estructura de datos que se utiliza para el entrenamiento del clasificador SVM queda conformada, por un lado, por valores de las distintas variables predictoras provenientes de registros geofísicos de pozos, propiedades petrofísicas o incluso atributos sísmicos, por otro lado, de los distintos valores de salida que, en este caso, serán cada una de las facies correspondientes a cada conjunto de variables predictoras que se haya seleccionado. La estructura de datos ya conformada se revisa y se depura para dejarla lista para que el clasificador SVM pueda realizar el entrenamiento. Una vez concluido, se obtiene un modelo de predicción, el cual se evalúa determinando las facies de vectores característicos que no le fueron dados a conocer al clasificador SVM durante el entrenamiento. De esta manera, solo resta a realizar la aplicación del modelo de predicción obtenido a todos los nuevos vectores característicos para realizar la predicción de sus facies.

2.5.4 Metodología del desarrollo del clasificador SVM

La metodología propuesta en la herramienta IMP-Predict se conforma de cuatro etapas (ver Figura 10):

- A. Exploración del conjunto de datos: esta etapa es de preparación en la cual se revisa el conjunto de datos que se va a utilizar para entrenar al clasificador SVM, se cargan todos los datos de entrenamiento, se grafican para el análisis de la variación dentro de los datos.

- B. Acondicionamiento del conjunto de datos: es la etapa que comprende la selección de las variables predictoras de entrada que se requieren para realizar la clasificación.
- C. Entrenamiento del clasificador SVM: en este apartado se utiliza el conjunto de datos acondicionado y depurado para crear el clasificador SVM, Los SVM definen un mapa de vectores característicos, que son puntos en un espacio multidimensional, los cuales son mapeados de modo que dichos vectores se dividan por una distancia que sea lo más amplia posible, porque un valor de C bajo hace la superficie de decisión suave y un valor alto trata de clasificar absolutamente todos los vectores de entrenamiento correctamente.
- D. Evaluación del clasificador SVM: en esta etapa se parte de la selección óptima de parámetros para el modelo de predicción con el fin de obtener resultados mejorados. Posteriormente, se aplica el modelo de predicción a datos ocultos, los cuales son los que no fueron utilizados para el entrenamiento con el fin de validar dicho modelo. Finalmente, se aplica el modelo de predicción a datos nuevos y se evalúan los resultados. Aquí se definen otras métricas también de utilidad conocidas como: precisión (*Precision*), sensibilidad (*Recall*) y F1 con el fin de brindar información puntual sobre el funcionamiento y desempeño del clasificador SVM. La precisión es la probabilidad de que, dado un resultado de clasificación para una muestra, la muestra pertenezca realmente a esa clase. La sensibilidad es la probabilidad de confirmar que una clasificación que se ha predicho sea la correcta. Por último, F1 es el cálculo de una función que combina los resultados de precisión y sensibilidad y puede tomarse como la medida de relevancia para conocer sobre el resultado de desempeño del clasificador SVM (Romero M, Diaz M, Del valle R, Albino J, 2018).

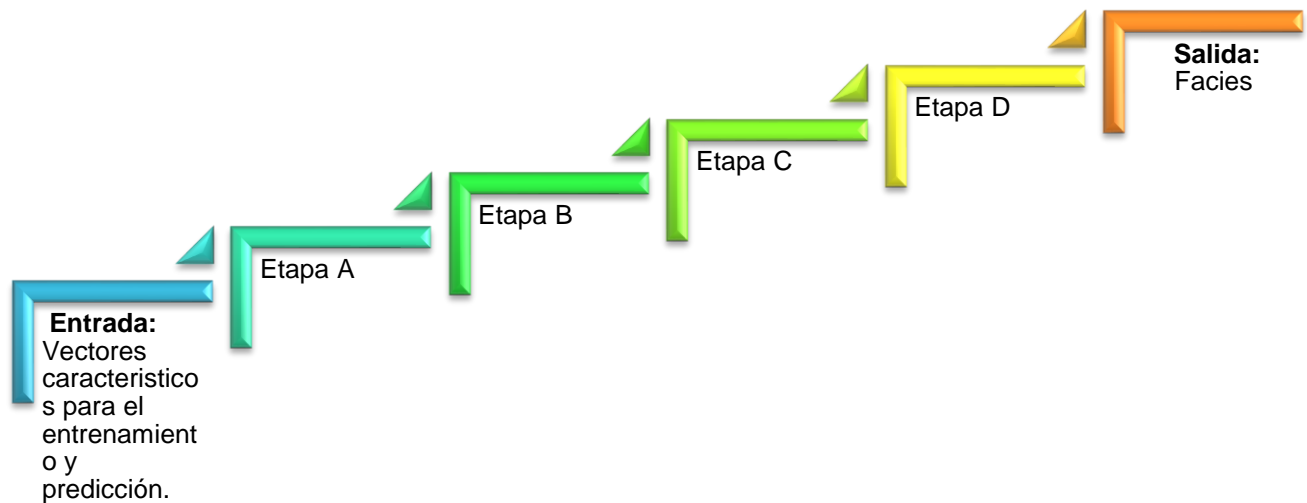


Figura 10.- Metodología para el desarrollo de un clasificador SVM.

2.5.5 parámetros del clasificador SVM: C & γ

Se consideran dos parámetros claves en un clasificador SVM, estos son: C y Gamma. Por un lado, el parámetro C es un factor de regularización que le indica al clasificador cuánto se desea evitar una mala clasificación de los vectores característicos utilizados para el entrenamiento. Un valor grande de C hará que el modelo de predicción intente clasificar correctamente más vectores característicos dentro del conjunto de datos de entrenamiento. Sin embargo, si C fuera demasiado grande, podría "compensar" los datos y al mismo tiempo no generalizar al clasificar nuevos datos. Por el contrario, si C fuera demasiado pequeño, el modelo podría no ser acertado para ajustar valores atípicos y entonces caería en un margen de error grande para el conjunto de vectores característicos.

Por otro lado, el algoritmo de aprendizaje, dentro del clasificador SVM, utiliza una función de base radial (Radial Basis Function o RBF) que permite calcular la distancia entre los vectores característicos. Dentro de esta función, existe un parámetro libre conocido como Gamma, el cual describe el tamaño de las funciones de base radial, que se interpreta como la distancia a la que se deben considerar dos vectores en el espacio. Un Gamma grande conduce a una alta tendencia y a modelos de predicción con varianza baja y viceversa.

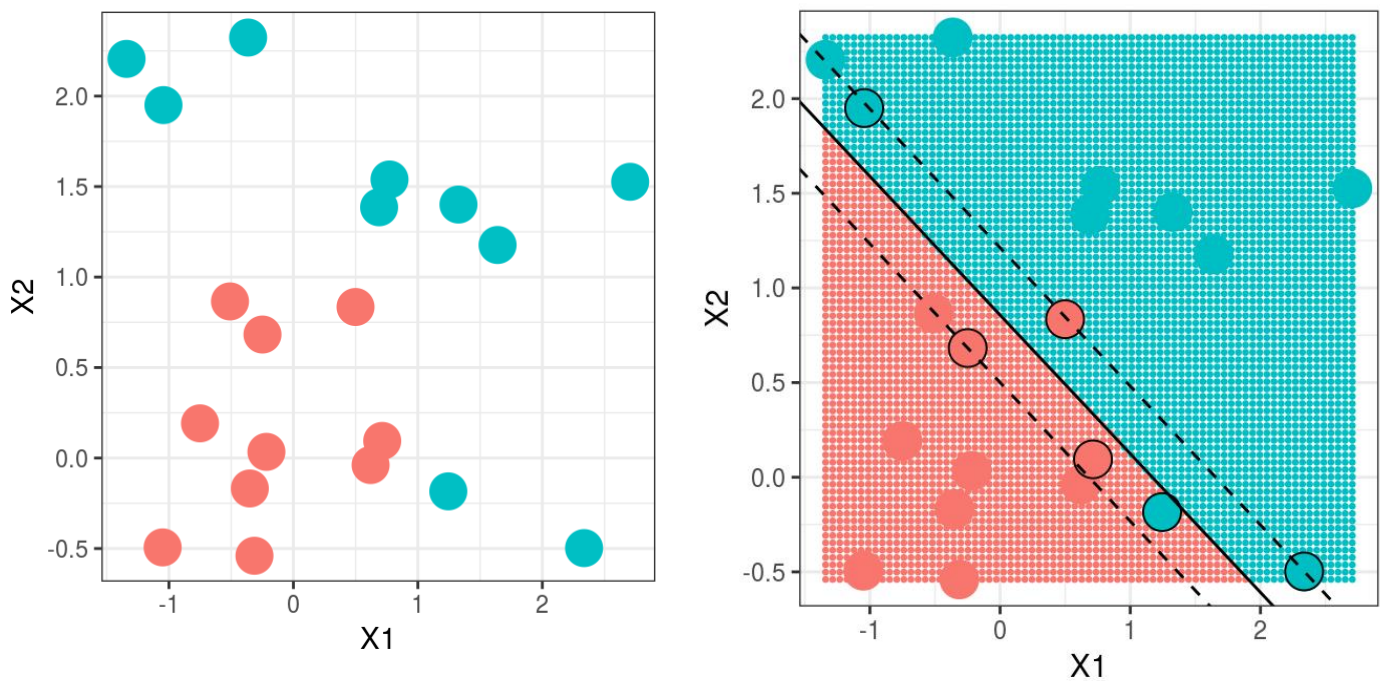


Figura 11.- Ejemplo gráfico de la clasificación realizada por el SVM de un conjunto de datos.

2.6 K-means Clustering en Python con Scikit-Learn

En este trabajo se realizará la identificación de las facies litológicas a partir de los registros de pozo y los datos de microestructura utilizando un enfoque basado en las redes neuronales de agrupación.

En el aprendizaje automático, los tipos de aprendizaje se clasifican en:

1. Aprendizaje supervisado
2. Aprendizaje no supervisado
3. Aprendizaje semi-supervisado

Los algoritmos que pertenecen a la familia de aprendizaje no supervisado no tienen variables para predecir vinculadas a los datos. En lugar de tener una salida, los datos solo tienen una entrada que serían múltiples variables que describen los datos. Aquí es donde entra el agrupamiento.

La agrupación por medio de clústeres es la tarea de agrupar un conjunto de objetos de manera que los objetos en el mismo grupo sean más similares entre sí que a los objetos en otros grupos. La similitud es una métrica que refleja la fuerza de la relación entre dos objetos de datos. La agrupación se utiliza principalmente para la minería de datos exploratoria. Tiene múltiples usos en muchos campos, como aprendizaje automático, reconocimiento de patrones, análisis de imágenes, recuperación de información, bioinformática, compresión de datos y gráficos por computadora.

Sin embargo, esta investigación mencionará el funcionamiento interno de K-means, una técnica de agrupamiento muy popular. El algoritmo permitirá abordar los conjuntos de datos sin etiquetar (es decir, los conjuntos de datos que no tienen etiquetas de clase) y extraer sus propias inferencias de ellos con facilidad.

2.6.1 Descripción del algoritmo de K-means

El algoritmo K-means, creado por MacQueen en 1967 es el algoritmo de clustering más conocido y utilizado ya que es de simple aplicación además de eficaz. Sigue un procedimiento de clasificación simple de un conjunto de objetos en un determinado número K de clústeres, K determinado a priori. El nombre de K-means viene porque representa cada uno de los clúster por la media (o media ponderada) de sus puntos, es decir, por su centroide. La representación mediante centroides tiene la ventaja de tener un significado gráfico y estadístico inmediato. Cada clúster, es caracterizado por su centro o centroide (ver Figura 10) que se encuentra en el centro o el medio de los elementos que componen el clúster- means que es traducido como K-medias.

El algoritmo del K-means se realiza en 4 etapas:

- Etapa 1 **inicialización**: una vez escogido el número de grupos k , se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.
- Etapa 2 **asignación objetos a los centroides**: cada objeto de los datos es asignado a su centroide más cercano. Para cada objeto x , el prototipo que se le asigna es el que es más próximo al objeto, según una medida de distancia (habitualmente la medida euclidiana).
- Etapa 3 **actualización centroides**: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.
- Etapa 4 **iteración** : Repetir las etapas 2 y 3 hasta que no se hagan más reasignaciones. Aunque el algoritmo termina siempre, no se garantiza el obtener la solución óptima. En efecto, el algoritmo es muy sensible a la elección aleatoria de los K centros iniciales. Esta es la razón por la que se utiliza el algoritmo del K-means numerosas veces sobre un mismo conjunto de datos para intentar minimizar este efecto, sabiendo que a centros iniciales lo más espaciados posibles dan mejores resultados.

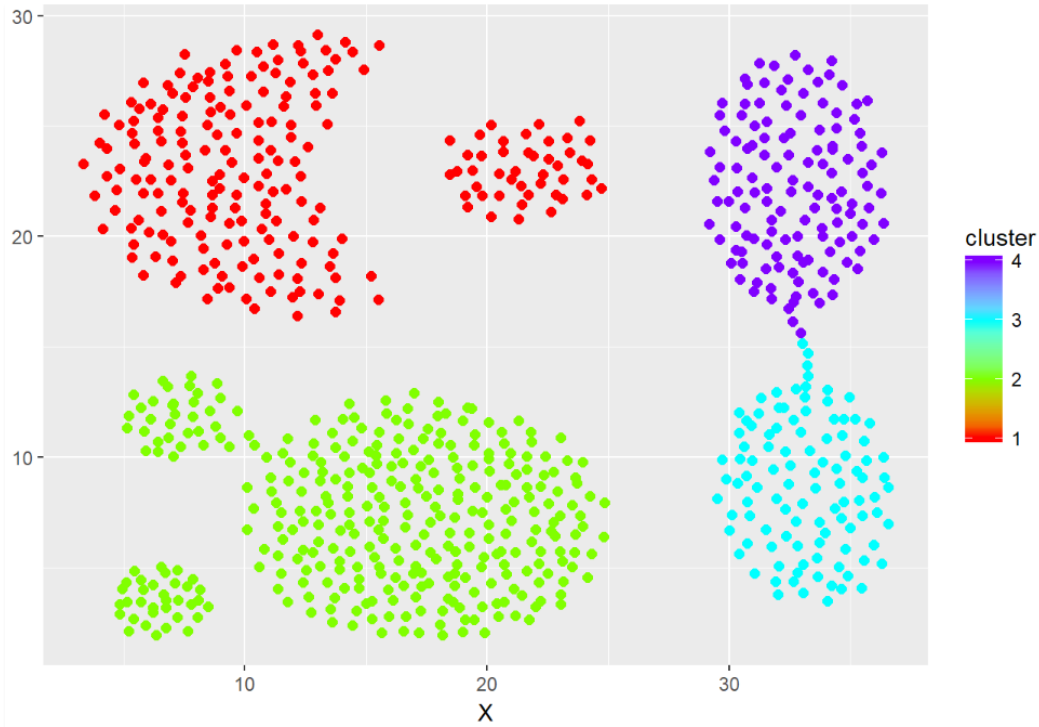


Figura 12. Clúster de datos con K=4 (Tomada de Sayak, P. 2018).

2.6.2 Ventajas y desventajas del método K-means

Este método trabaja con el conjunto completo de referencia y su objetivo es reducir el número de cálculos, eliminando aquellos que se consideran “inútiles” o “innecesarios”

La utilización de este método no representa ningún peligro de pérdida de generalidad ya que está orientado, únicamente a reducir el número de cálculos. No obstante, hay que considerar que este método tiene asociado un costo adicional a la búsqueda, el costo del preprocesamiento, como el asociado a la selección del conjunto reducido o la construcción del árbol de búsqueda. En consecuencia, para determinadas aplicaciones se ha de considerar que el costo de preprocesamiento puede ser alto .

Ventajas:

- El costo del aprendizaje es nulo.
- No se necesita hacer ninguna suposición sobre los conceptos por aprender.
- Aprende de conceptos complejos usando funciones sencillas como aproximaciones locales.
- Desarrolla el mecanismo para predecir un valor continuo de regresión.
- Es muy tolerante al ruido.

Desventajas:

- Requiere que se especifique previamente el número de grupos (k).
- Es necesario realizar en sucesivas ocasiones el algoritmo para así tener el resultado más óptimo posible. Tiene la necesidad de inicializar el número de prototipos al principio de la ejecución. Esto perjudica la eficacia del algoritmo ya que en la práctica, no se conoce a priori el número de clúster final.
- k-means es susceptible a valores extremos porque distorsionan la distribución de los datos (Sayak,P. 2018).

2.7 Resumen del capítulo

En este capítulo se presentaron los temas relacionados con la propuesta de investigación. Se describieron los aspectos teóricos y técnicos de los métodos a desarrollar, los distintos temas que abordan la metodología, los alcances y limitaciones conocidas previamente. Se mostraron los registros geofísicos que se utilizarán, así como la información petrofísica de las rocas carbonatadas que existen en los datos, de igual manera se mostró los aspectos teóricos dentro de las herramientas de aprendizaje automático, del cual comprende dos tipos de aprendizaje, supervisado y no supervisado. En el siguiente capítulo se mostrará la propuesta de clasificación desarrollada, donde se incluyen los aspectos preliminares requeridos para el seguimiento de las secciones descritas, la metodología completa dividida en cinco secciones principales y los aspectos prácticos de ajustes y calibración.

Capítulo 3. Metodología

3.1 Introducción

Una buena representación de facies es de gran utilidad para que el intérprete pueda definir qué zonas son de interés y cuáles deben de ser descartadas. En este capítulo se presentará el proceso de cálculo desarrollado para llevar a cabo una alternativa de clasificación de facies. El proceso de cálculo involucra un conjunto de scripts basados en la librería de Scikit Learn, el cual tiene como entrada un conjunto de datos y consigue realizar una especie particular de clasificación, que describe cada una de las facies. La idea general de la propuesta se basa en una búsqueda controlada de elementos con características similares. Se mostrará que este procedimiento es adecuado para encontrar diversas facies, manejando las propiedades petrofísicas y los registros de pozos que suelen encontrarse por procesos directos ya antes mencionados, previamente se describirán los elementos técnicos de los códigos.

De igual manera en este capítulo se dará a conocer el tratamiento de los datos obtenidos por Instituto Mexicano del petróleo y la preparación de los datos para procesamiento en la herramienta de inteligencia artificial.

3.2 Metodología del trabajo

En esta sección se presenta el flujo de trabajo principal desarrollado en esta tesis. Se describirá los tipos de datos que se utilizaran y posteriormente cada uno de los módulos que lo conforman. Se describe también la manera en la que se implementó el algoritmo principal. Finalmente, se muestran algunos de los resultados que se obtuvieron, aplicando el algoritmo descrito en esta sección. En la figura 13 se muestra el esquema general del flujo de trabajo que consiste en cinco etapas principales de procesamiento, que se irán describiendo en las secciones subsecuentes. De manera general, se trata de un proceso de adquisición, calibración, procesamiento y visualización de resultados.

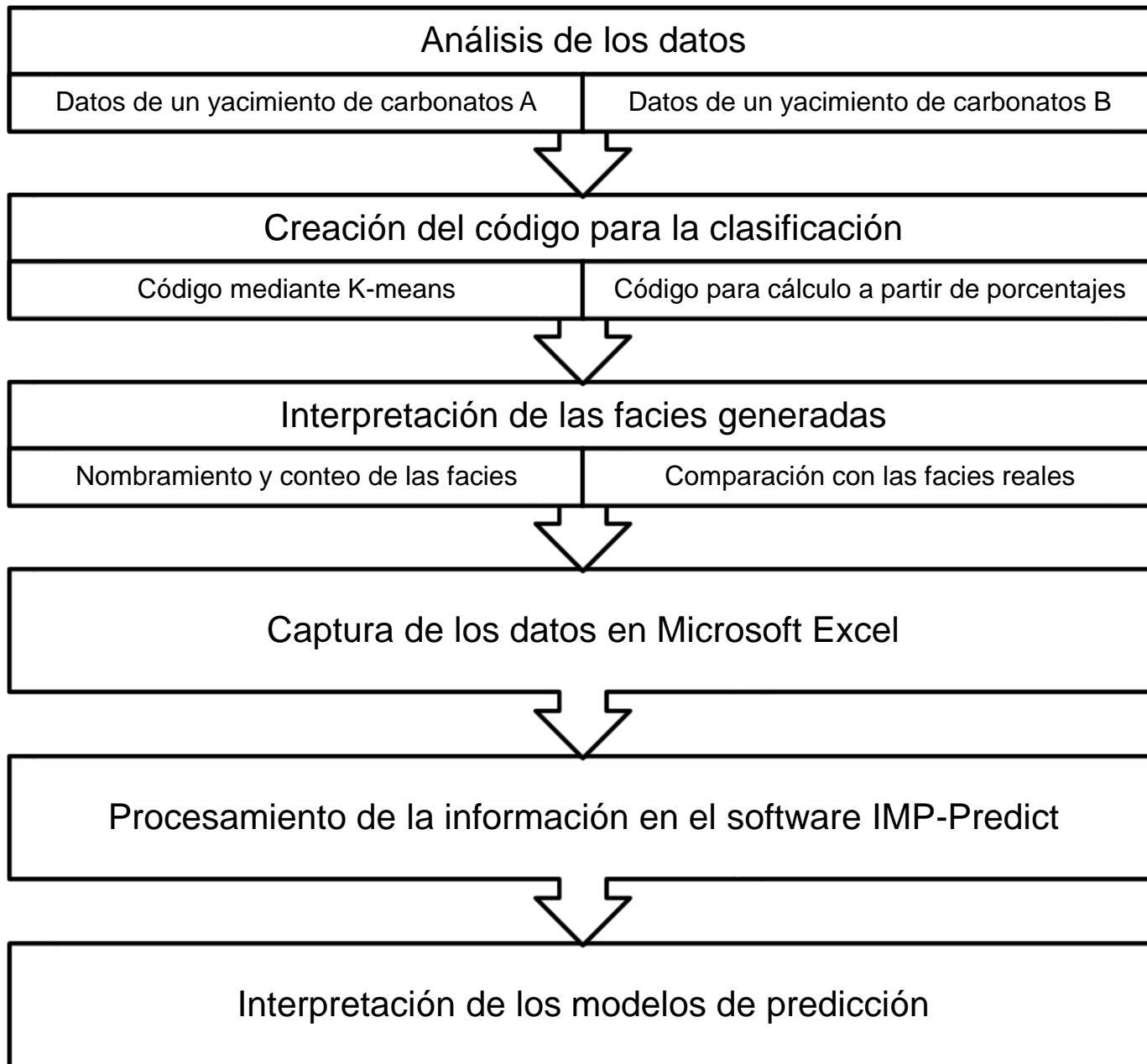


Figura 13.- Esquema general del flujo de trabajo.

3.3 Base de datos

Las bases de datos que serán procesadas consiste en dos hojas de Excel las cuales contienen información de registros de pozos y de microestructura con la información sintética de su profundidad de un yacimiento carbonatado (ver tabla 4,5), las cuales fueron proporcionadas por un especialista del Instituto Mexicano del Petróleo.

Nombre del registro	Máximo	Profundidad	Mínimo	Profundidad
A.Matrix.Porosity “ Matriz de porosidad ” (M PORO)	0.084792	288	0.0050297	1
A.Matrix.Dolomite_C “ Matriz de dolomitas ” (MDOL)	0.99999	815	0.13494	653
A.P1.Concentration “ Concentración de Vúgulos ” (CVUGU)	0.21684	293	0.0078198	552
A.P2.Concentration “ Concentración de fracturas ” (CFRAC)	0.015179	378	0.0027657	1073
Shale.Concentration “ Concentración de lutita ” (SC)	0.96478	329	0.062961	343

Tabla 4. Información de las propiedades petrofísica del yacimiento de carbonatos A.

Nombre del registro	Máximo	Profundidad	Mínimo	Profundidad
GR “ Rayos gamma ”	90.597	3032.6	25.1239	3317.8
RT “ Registro resistivo ”	103.1891	3081.1	0.4758	3161.9
NPHI “ Porosidad del registro neutrón ”	0.5093	3067.8	0.0554	3340.6
RHOB “ Densidad de formación ”	2.6031	3324.3	1.8506	3032.6
DTCO “ Tiempo de transito computarizado ”	124.5048	3082.1	66.138	3339.9

Tabla 5. Información de los RGF de los datos del yacimiento de carbonatos B.

3.4 Creación del código para la clasificación de facies

La clasificación de las facies en un cierto intervalo se aborda desde el enfoque de clasificación no supervisada, usando como herramienta K-means con el lenguaje de programación Python, específicamente desarrollado en *PYCHARM* (Jetbrains, 2000). Como se mencionó en el capítulo anterior es eficaz en las ciencias aplicadas, para determinación de soluciones y predicción del comportamiento de las mismas, sin recurrir a la interpretación directa de núcleos, petrografía, diagénesis o ambientes de depósito. Para realizar la clasificación se creó un programa que sirve como herramienta clasificadora de facies, utilizando las variables petrofísicas o registros geofísicos de pozos, así como las clases que se consideren necesarias. El programa realiza iteraciones con las variables que tú le indiques, lo cual asocia el agrupamiento de la información con el valor medio más cercanos en un espacio n-dimensional (Figura 14).

Por otro lado tenemos otro código que de la misma forma fue desarrollado en Python con el propósito de determinar facies de manera cuantitativa por medio de porcentajes que le otorgas a las variables petrofísicas, dicho código solo fue utilizado con la información de petrofísicas del yacimiento de carbonatos de la prueba 1. La forma de ejecución del código es similar al de k-means (Figura 15), solo que tú determinas el porcentaje que se le dará a cada variable, esto dependerá de tu análisis previo de la información. En lo que respecta a esta tesis, el análisis de los porcentajes se llevó a cabo por medio de un análisis estadístico de cada curva de los registros petrofísicos, obtenido el promedio total de la información de cada variable y observando su comportamiento en gráficos de dispersión generados en Microsoft Excel. Una vez realizado el análisis se procede a delimitar el porcentaje y realizar las iteraciones de la información en el código, para crear la clasificación en un cierto intervalo del yacimiento para después poder procesarla en la herramienta IMP-Predict V 2.0.

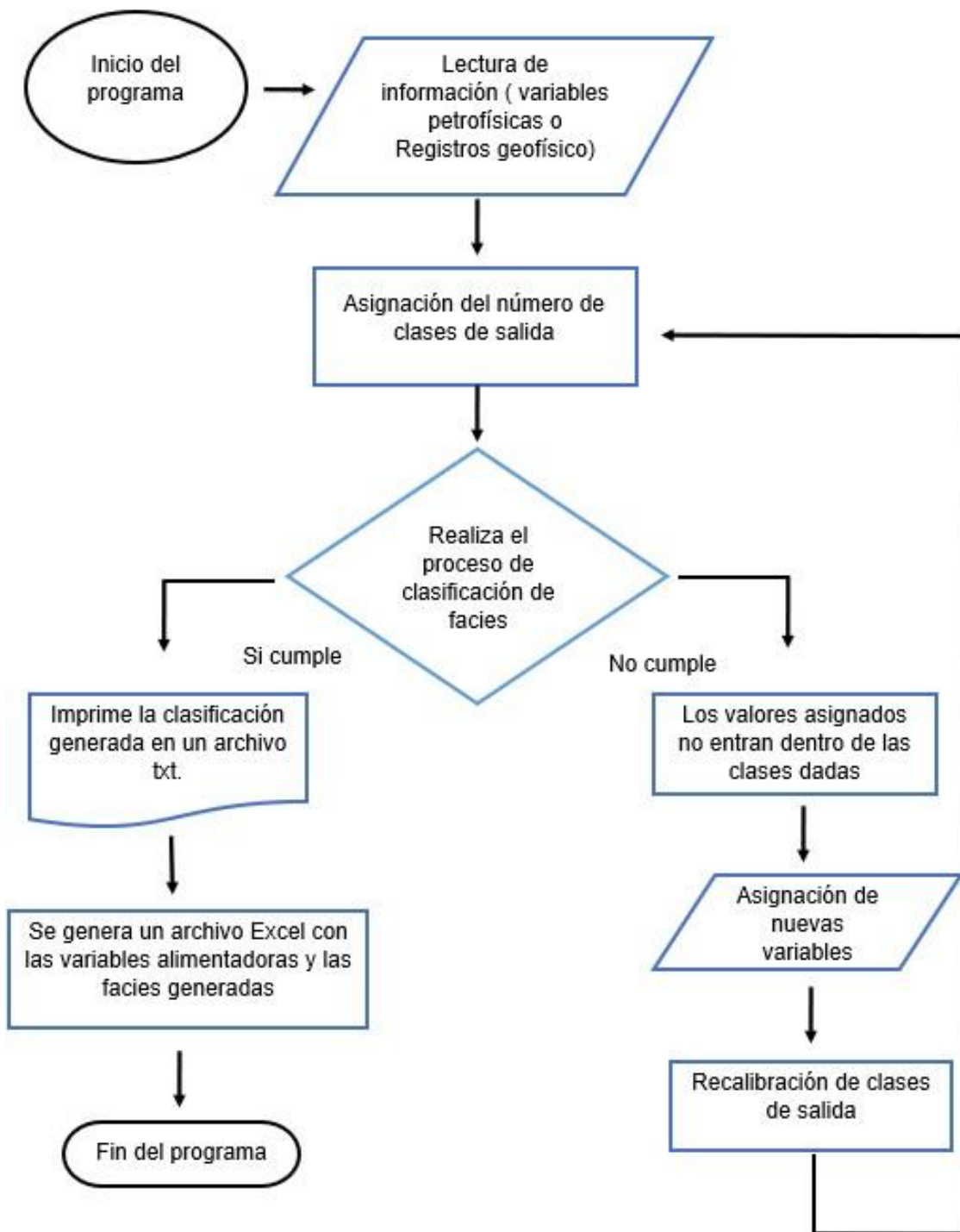


Figura 14. – Diagrama de flujo del método de clasificación de facies por medio del algoritmo de K-means.

Los diagramas de flujo son una herramienta útil para comprender la lógica del programa utilizado, estos nos muestran de manera esquemática cuales son los datos que ingresamos en el programa y los resultados que obtendremos, mediante el uso de flechas se observa el flujo que tendrán los datos, mediante rectángulos se representan sentencias utilizadas y usando rombos se representan disyuntivas.

En la parte superior del diagrama podemos observar un ovalo con la leyenda “inicio del programa” (Figura 14), en esta parte se infiere el inicio del código, posteriormente se define las variables a utilizar y el nombre que tomarán dentro del programa. Las variables utilizadas son de tipo numérico, específicamente valores enteros y decimales(Int, Float). Para este caso son los registros petrofísicos o geofísicos que se consideren de importancia para obtener la clasificación correcta.

Una vez declaradas las variables, se le pide al usuario del programa que asigne el número de clases, así como también se requiere que se ingrese el archivo de Excel o txt. Del cual tomara esas variables, una vez dado todo estos datos por el usuario el código en Python empezará a generar la clasificación de acuerdo a las declaraciones previas, si la información otorgada no cumple con las especificaciones necesarias se necesitará regresar al apartado anterior para realizar la calibración de los datos con el fin de obtener el correcto flujo de trabajo. Finalmente se registran los resultados de ambos programas en un archivo .txt y se imprimirán en un archivo de Microsoft Excel, donde viene dada la clasificación de las información con sus respectivas variables que se tomaron para el procesamiento.

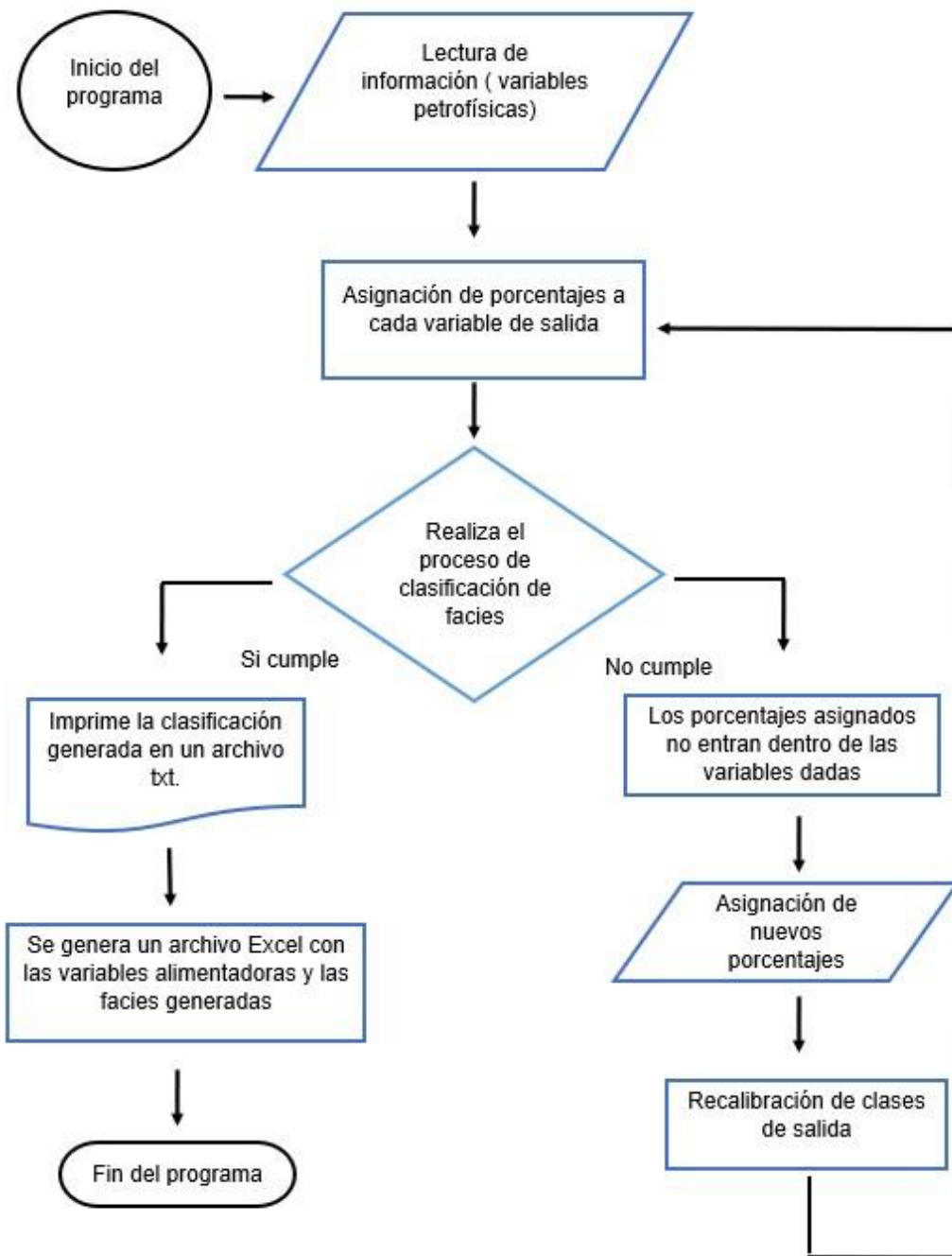


Figura 15. – Diagrama de flujo del método de clasificación de facies por medio de asignación de porcentajes.

3.5 Descripción de facies generadas

Una vez obtenida la clasificación por medio de clúster o porcentajes para cada caso, se procede a desarrollar una interpretación que permita tener una visualización general del comportamiento del subsuelo. Con el desarrollo de estas clasificaciones se pretende determinar si la respuesta evaluada tiene un sentido geofísico lógico.

En esta sección se analizarán los resultados generados a partir de las iteraciones del programa propuesto por el algoritmo K-means. Se han hecho varias iteraciones con el fin de obtener una mejor clasificación y para el caso del programa por medio de porcentajes se han hecho modificaciones en los pesos que se le da a cada variable

Por otro lado con los datos de un yacimiento carbonatado A, donde no se tiene información directa de su marco geológico y de sus facies reales para la comparación. Se obtienen dos clasificaciones, por medio del algoritmo K-means y por asignación de porcentajes a cada parámetro.

En las clasificaciones dadas se realiza el nombramiento de facies conforme a la información que se utilizó para generar las facies. En el caso de esta investigación solo se tiene como información que son datos de un yacimiento carbonatado con su profundidad sintética en la cual se presentan concentraciones de rocas calizas, lutitas y dolomitas, con porosidades; primaria de matriz y porosidad secundaria conformada por vóculos y fracturas.

La tabla 6 presentan las facies con sus etiquetas dadas por el algoritmo de K-means lo cual se le asignó como variables los datos petrofísicos (Matriz de porosidad, concentración de dolomitas, concentración de vóculos, de fracturas y concentración de lutita) en el caso de la tabla 7 se utilizaron los datos de registro geofísicos de pozos (GR,RT,NPHI,RHOB Y DTCO) que son los registros convencionales que se utilizaron para el desarrollo de facies reales, en ambas clasificaciones se manejaron cuatro facies, el número fue propuesto por el doctor del IMP, debido a que la información manejada es escasa.

Facies	Cantidad de datos
Facies 1 (C) “caliza con alta concentraciones de lutita con baja concentración de vgulos, concentracin alta de fracturas y concentracin media de porosidad primaria ”	281
Facies 2 (C-D) “caliza dolomitizada con alta concentraciones de lutita ,media concentracin de vgulos, media concentracin de fracturas y media porosidad de matriz”	327
Facies 3 (D-C) “dolomita calcrea, media concentracin de vgulos, concentracin media de fracturas y baja porosidad de matriz con poca impregnacin de lutita”	57
Facies 4 (D) “dolomita media concentracin de vgulos baja concentracin fracturas, y baja porosidad de matriz, con poca impregnacin de lutita”	494

Tabla 6. Clasificacin generada por medio de clster (K-means) con las propiedades petrofsica del yacimiento de carbonatos A.

Facies	Cantidad de datos
Facies 1 (C) “caliza”	1066
Facies 2 (C-D) “caliza con dolomita”	1696
Facies 3 (D-C) “ doloma calcrea“	1208
Facies 4 (D) “dolomita”	62

Tabla 7. Clasificacin de las facies reales del yacimiento de carbonatos B, con las variables de registro geofsicos de pozos.

En el caso dos, donde se asignan porcentajes a cada variable de los datos de microestructura, lo cual producen facies de manera interpretativa tratando de localizar el tipo de roca presente como sus porosidades primarias y secundarias. Los porcentajes fueron dados bajo un análisis de la información y con la ayuda del personal del IMP, en los siguientes puntos se explica a qué se refiere cada porcentaje en valor interpretativo de la información del yacimiento carbonatado.

- En matriz de dolomitas (MDOL): con el fin de obtener el tipo de roca predominante en la zona. Se manejan 4 tipos de pesos que van de 25% -50%-75%-100%, donde se concreta que si los datos entran en el rango de 0- 25% es una roca caliza, si está en el rango de 25 – 50% es una caliza dolomitizada, si entra en el rango 50 – 75% es una dolomita calcárea y si está en el rango de 75 – 100% es una dolomita pura.
- En concentración de lutitas (C.LUTITA): los rangos manejado fueron (5%- 30 % - 100%), este peso se debió a que el volumen de información es pequeño, en el cual se interpreta que la roca al tener 0-5 % no tiene presencia de lutita , del 5- 30% hay poca presencia de lutita y del 30 – 100% hay impregnación de lutita.
- En porosidad de matriz (M PORO): se utilizaron porcentajes que van de (15% - 50%-100%). Esto me permite definir en qué facies existe este tipo de porosidad primaria, haciendo énfasis en que si tienen un porcentaje menor del < 15%, no se presenta dicha porosidad, si tiene más del > 50% su concentración es media y si entra en el rango de 50% a 100% tiene una alta concentración de porosidad primaria.
- En el caso de la concentración de vórgulos (CVUG) y concentración de fracturas (CFRAC), solo se le dieron tres denominaciones, para indicar si las porosidades secundarias están presentes en la roca y su proporción. los pesos que se manejaron fueron de 0- 25% - 50%- 100% de cada variable, lo cual indica, al entrar en el rango de 0- 25% hay poca presencia de este tipo de porosidad, 25-50% hay media concentración de porosidad secundaria y si va de 50-100% hay una alta concentración de porosidad secundaria.

Facies	Cantidad de datos
Facies 1 (C) “caliza con alta impregnación de lutita, baja concentración de vúgulos, concentración media de fracturas y media concentración de porosidad primaria ”	51
Facies 2(C-D) “caliza dolomitizada sin presencia de lutita media concentración de vúgulos, baja concentración de fracturas y sin porosidad de matriz”	103
Facies 3 (D-C) “dolomita con media concentración de vúgulos, concentración media de fracturas y alta concentración de porosidad primaria con contenido medio de lutita”	670
Facies 4 (D) “dolomita alta concentración fracturas, vúgulos y baja porosidad de matriz , sin impregnación de lutita.	335

Tabla 8. Clasificación generada por medio de porcentajes de las propiedades petrofísicas.

3.6 Preparación de datos

Para el desarrollo de la experimentación se cuenta con 2 yacimientos carbonatados (petrofísicos y registro geofísicos) con sus respectivas facies dadas por lo diferentes métodos. Para realizar el procesamiento en la herramienta se deben seguir los siguientes pasos:

1. Generar un archivo de Excel con las 3 clasificaciones de facies diferentes con sus respectivas variables que las generaron.
2. Definir los intervalos que se usarán conforme a la profundidad presentada (Tabla 9 y 10).
3. Acomodar los datos conforme el software IMP-Predict 2.0 los puede procesar, en el cual se debe tener 3 archivos txt ; uno de entrenamiento nombrado (E), otro de predicción nombrado (P) y el ultimo con los datos reales (ALL).
4. Nombrar por cada clasificación una prueba para el modelo de predicción(Tabla 11).

Intervalo del pozo	Profundidad que comprende
Intervalo 1	1 - 280 m
Intervalo 2	281 - 860 m
Intervalo 3	861 – 1159 m

Tabla 9. Intervalos definidos en los datos de Excel para la pruebas 1 y 2 a procesar.

Intervalo del pozo	Profundidad que comprende
Intervalo 1	2998.7- 3100 m
Intervalo 2	3100.1 - 3230 m
Intervalo 3	3230.1 – 3404.5 m

Tabla 10. Intervalos definidos en los datos de Excel para la prueba 3 a procesar.

Pruebas
Prueba 1: datos del yacimiento A de carbonatos con las información petrofísica y sus facies generadas por K-means.
Prueba 2: datos del yacimiento A de carbonatos con las información petrofísica y sus facies generadas por asignación porcentajes
Prueba 3: datos del yacimiento de carbonato B con información de registros geofísico de pozo y sus facies reales.

Tabla 11. Pruebas para el procesamiento en IMP-Predict.

3.8 Resumen del capítulo

En este capítulo se presentaron los programas a utilizar en este trabajo, como lo son el algoritmo de k-means, porcentajes y la herramienta IMP-Predict v 2.0. Se describieron algunos conceptos preliminares que el lector debe de considerar para un seguimiento del capítulo. Se describe el flujo de trabajo de la metodología, que consiste en cinco etapas principales: Adquisición de datos, análisis, creación de los scripts de clasificación de facies, interpretación de las facies y procesamiento en IMP-Predict. Se mostraron los resultados obtenidos en la clasificación. También se mostró una métrica cuantitativa que sirve para evaluar el desempeño del algoritmo. En el siguiente capítulo se presentan los diversos casos de estudio que demuestran su desempeño. Se analizan tres pruebas específicas. Así mismo, se demostrará que el algoritmo presentado en este trabajo de investigación es capaz de obtener una clasificación congruente con la interpretación de los modelos de predicción.

Capítulo 4. Resultados: análisis de los modelos de predicción

El desarrollo de las facies es de gran importancia en el proceso de investigación e interpretación de registros, ya que con esto podemos obtener la litología presente de una manera fácil, una vez realizada la interpretación en un registro, se pueden seguir utilizando los mismos parámetros del registro anterior para los pozos adyacentes.

El objetivo de este capítulo es aplicar y demostrar la utilidad de las facies generadas con los métodos propuestos anteriormente. La interpretación de las facies se lleva a cabo en 2 bases de datos de origen de rocas carbonatadas. Todos los métodos fueron realizados individualmente y comparados entre ellos para comprobar la interpretación y resultados. Para realizar los modelos de predicción de las facies se seleccionaron las curvas necesarias para su interpretación, en el caso del yacimiento de carbonatos B, se utilizaron GR, RT, NPHI, RHOB y DTCO, que forman parte de un conjunto de registros de pozos. De igual manera se realizó la predicción con las facies generadas por medio de los datos de petrofísicos, las curvas que se utilizaron fueron MDOLO, CVUGU, CFRAC, MPORO Y C.LUTITA, en ambos entrenamientos se probaron todos los parámetro C y GAMMA para argumentar la hipótesis de cómo tener una mejor predicción a partir de ellos.

4.1 Selección de los parámetros óptimos para el modelo de predicción

Durante la etapa de procesamiento detallada en el capítulo anterior, el clasificador se ha construido con los parámetros por defecto ($c = 500$, $\gamma = .07$). Sin embargo, es posible obtener mejores resultados de clasificación manipulando algunas opciones de sus parámetros. Para ello, se lleva a cabo una validación cruzada con el fin de realizar la selección de los parámetros adecuados para el modelo de predicción. Con el fin de evaluar el comportamiento del modelo de predicción, utilizando diferentes valores para C y Gamma, se define una función con dos ciclos anidados que genera gráficas de "validación cruzada" (cross validation o CV) a través de las cuales se puede visualizar el error para cada posible combinación de valores en los intervalos indicados. Por lo que la

precisión del modelo de predicción puede ser mejorada encontrando una combinación adecuada de los valores de los parámetros C y Gamma. La elección de C y Gamma se lleva a cabo bajo el análisis de la curva azul, conocida como la curva de cross validation, en dicha curva se busca que los valores se ajusten lo más alto posible de la gráfica como se muestra en la figura 16, este análisis se llevó a cabo en las 3 pruebas que se entrenaron en la herramienta para encontrar el menor error posible, lo cual da la comprobación de manera experimental de otra de mis hipótesis establecidas

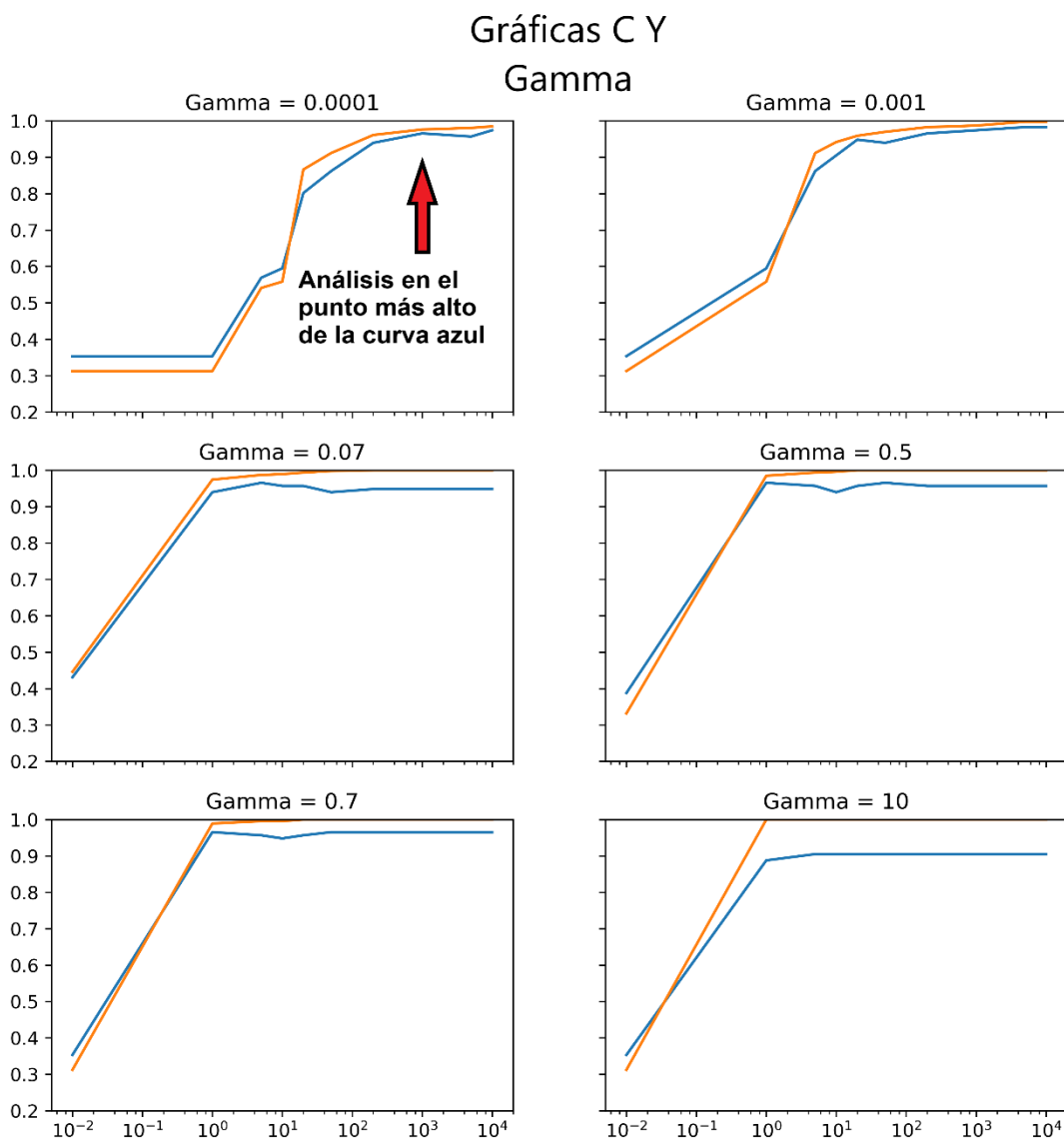


Figura 16. – Graficas C y Gamma del modelo de predicción de los datos carbonatados.

Intervalo de entrenamiento	Intervalo oculto	Intervalo a predecir
Intervalo 2	Intervalo 1	Intervalo 3

Tabla 12. Plan de entrenamiento para el procesamiento en IMP-Predict con las 3 pruebas de los yacimientos de rocas carbonatadas.

Facies	Simbología	Color
1 “caliza”	C	Rojo
2 “caliza con dolomita”	C-D	verde
3 “dolomía calcárea”	D-C	amarillo
4 “dolomita”	D	Azul

Tabla 13. Definición de facies, simbología y colores.

4.2 Modelos de predicción de la prueba 1: Datos Carbonatados.

Para los datos sintéticos se utilizaron las facies generadas por el método de clasificación K -means, donde se realizó el plan de entrenamiento como se detalla en la tabla 12. Se buscó por medio de comprobación de los parámetros C y Gamma, el modelo de predicción con menor margen de error. Por motivos de simplicidad solo se mostrarán los resultados óptimos generados por la herramienta.

4.2.1 Descripción del intervalo oculto

La herramienta utiliza un intervalo oculto realizando una predicción para efectos de calibración y así obtener el resultado de la predicción final. En el intervalo oculto se observa una predicción buena, donde existe una intercalación de facies 1,3 y 4:

- 0 - 100 m, se observa la facie 4 denotada en color azul, donde existe dolomita con media concentración de vugulos, baja concentración fracturas y baja porosidad de matriz, con poca impregnación de lutita.
- 100-110 m, en este intervalo pequeño se puede notar la facie 2, denotada de color verde.
- 220-240 m, se observa la facie 3 denotada de color amarillo.
- 240- 270 m se observa la predominancia de la facie 1, presentada de color rojo.

De manera objetiva se concluye que la rocas predominantes en esta profundidad es la dolomita con presencia de caliza o dolomitas calcáreas, existe una baja concentración de porosidad vugular, de fracturas y una concentración media de porosidad de matriz, se observa media impregnación de lutita en la roca, se muestra una intercalación de la facies 1, 3 que tienen una alta concentración de porosidad de matriz y porosidad secundaria de fracturas, en todo este intervalo se tiene media a baja concentraciones de lutita.

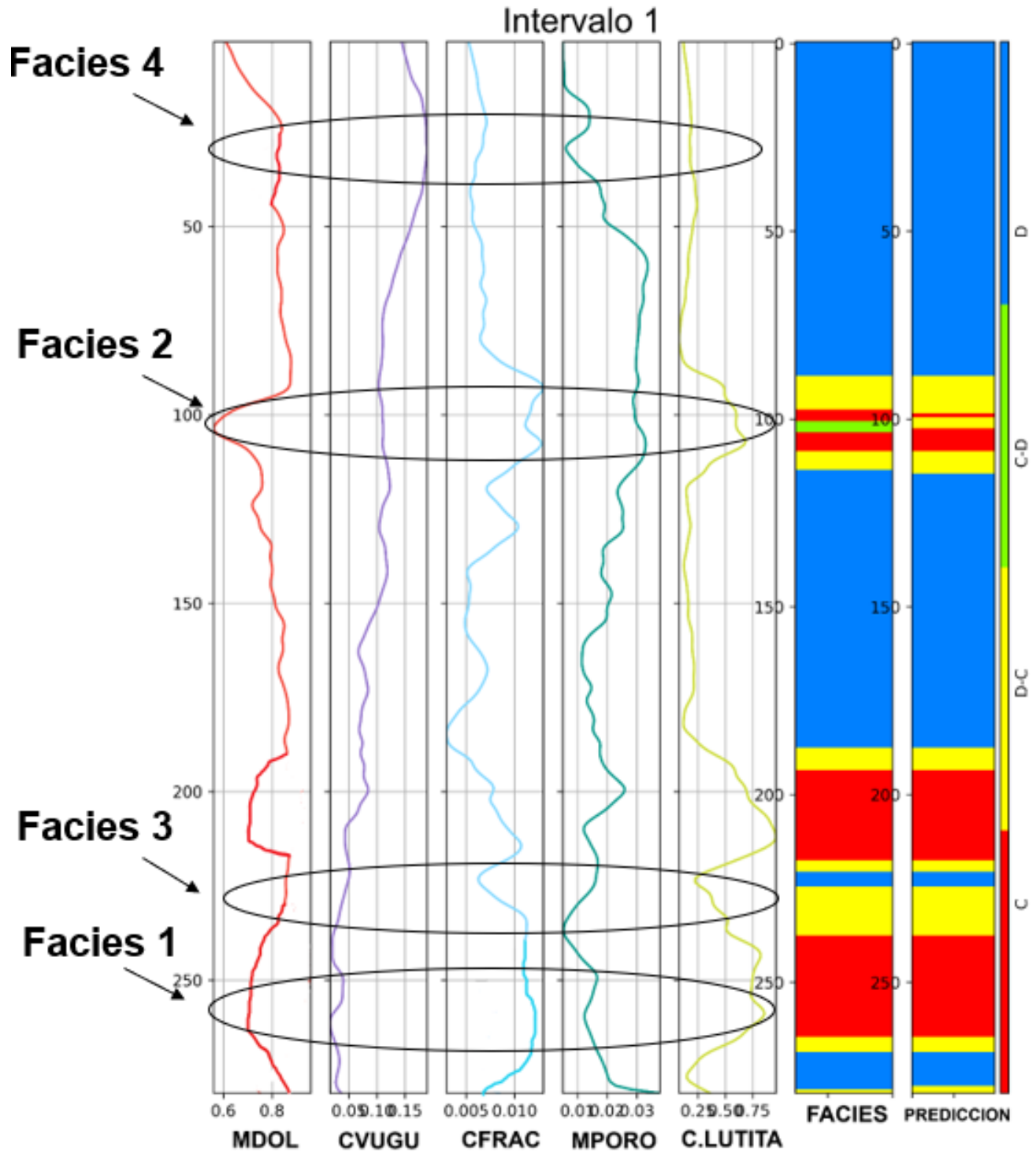


Figura 17. – Resultados de la predicción del intervalo oculto (int1).

4.2.2 Descripción de la predicción final dada en el intervalo 3

El resultado preliminar obtenido, tuvo un acierto de 94.98% (Figura 18) con un $C = 1000$, $\gamma = .001$. Se concluye que la herramienta creó modelos de predicción buenos comparados con las facies reales que se le dieron a iterar. Se observa intercalación de facies 3,4,1 mostrando que en este intervalo del yacimiento no hay presencia de la facies 2.

- 1050 – 1200 m, hay mayor concentración de dolomita con media concentración de vugulos baja concentración fracturas y baja porosidad de matriz, con poca impregnación de lutita
- 1030-1040 m, hay presencia de caliza con alta concentraciones de lutita baja concentración de vugulos, concentración alta de fracturas y concentración media de porosidad primaria
- 861 – 920 m, existe una alta concentración de porosidad vugular, baja concentraciones de porosidad por fracturas y de matriz representadas por las facies 4 y 3.
- En el último intervalo de profundidad 1020 – 1160 m, se observa un alta concentración de dolomita, dolomita calcárea, con concentraciones media de porosidad vugular, baja concentraciones de porosidad por fracturas y de matriz de porosidad, con poca impregnación de lutita, denotados por el color amarillo y azul, correspondientes a las facies 3-4.

se comprueba el comportamiento y la congruencia de las facies generadas con las curvas de las variables petrofísicas que se le dieron para el procesamiento.

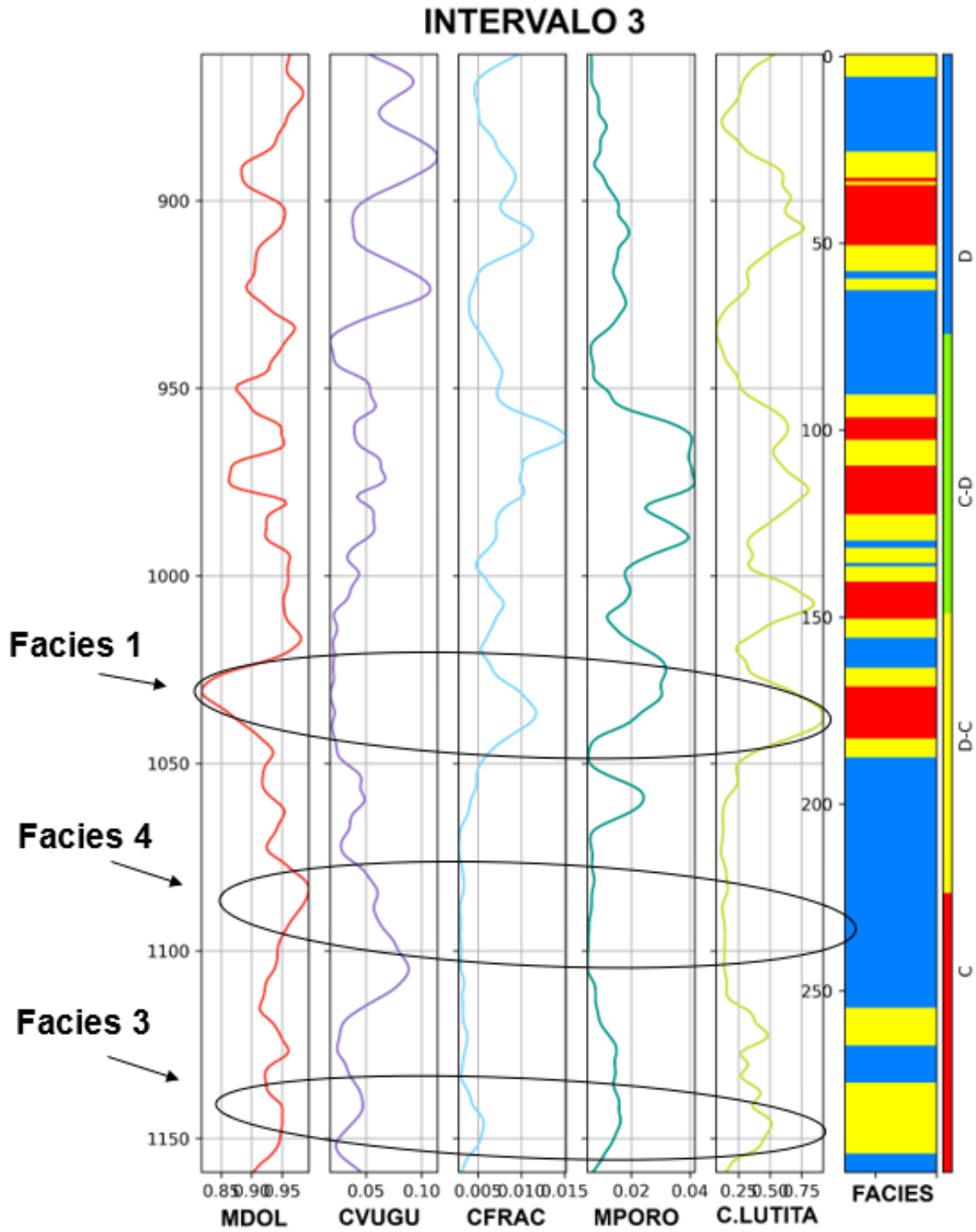


Figura 18. – Resultados de la predicción del intervalo 3.

4.3 Modelos de predicción Prueba 2 : Facies generadas por asignación de porcentajes

Una parte importante del desarrollo de este trabajo, se basa en el hecho de proponer una clasificación de facies litológicas por medio de la caracterización petrofísica de las rocas carbonatadas. Asumir lo anterior significa que por medio de la guía de un experto en la materia, con el previo análisis de las curvas y de los datos, se desarrollan porcentajes para cada información de microestructura, dando como resultado facies que nos informan el tipo de roca con su contenido de porosidad. Tal como se detalla en el capítulo 3, las facies generadas son descritas en la Tabla 8. A continuación se darán los modelos de predicción otorgados por la herramienta, se hará énfasis en el modelo generado con el mejor arreglo de C, γ . De igual manera como la prueba anterior, se utilizó el mismo plan de entrenamiento que consta de 3 intervalos, detallado en la tabla 12.

4.3.1 Descripción de la predicción en el Intervalo oculto (Intervalo 1)

Como se ha hecho mención antes, la herramienta solicita un intervalo o un pozo, para realizar una predicción, este intervalo se utiliza para efectos de calibración y así trabajar en conjunto con el intervalo de entrenamiento para otorgarnos la predicción final requerida. En la figura 19 se observa que el intervalo 1 con profundidad de 281-860 m, consta de una buena predicción. De manera descriptiva se observa que la facies predominante es la 4, la cual describe una litología conformada de roca dolomita con alta concentración de vóculos, concentración media de fracturas y porosidad de matriz sin lutita (color azul). Se observan intercalaciones de caliza dolomitizada (facies 2) y poca presencia de la facies 1 conformada de caliza con baja concentración de vóculos y alta de porosidad conectada de fracturas denotadas en rojo.

En el intervalo de 300- 600 m se tiene una intercalación de las 4 facies , demostrando la litología presente en el yacimiento carbonatado.

En el intervalo de profundidad de 100- 280 m, sigue un comportamiento similar donde las facies predominante es la 3 (D-C), la cual nos interpreta que existe una media a baja

concentración de porosidad vugular y una alta a media concentración de porosidad de fracturas y matriz.

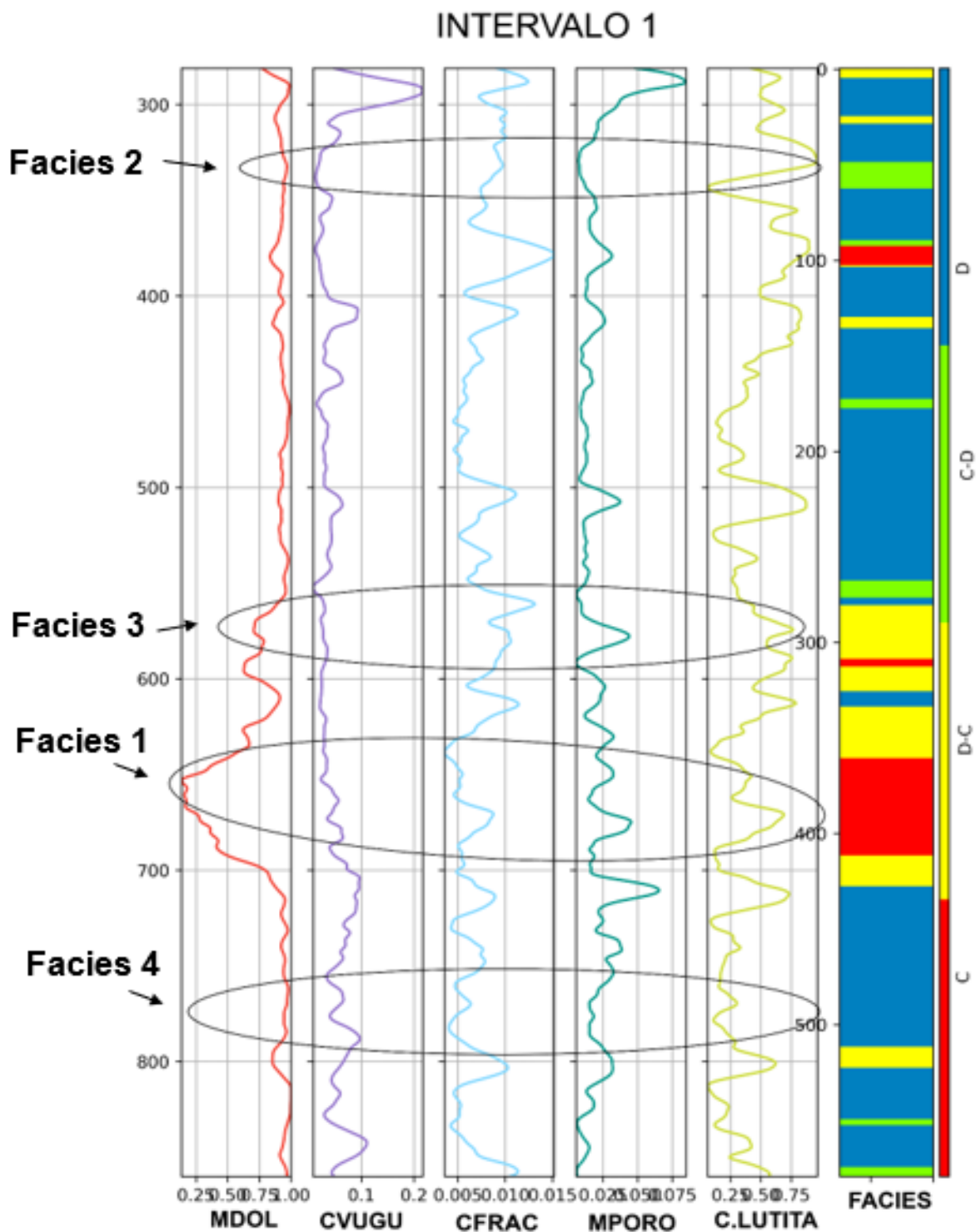


Figura 19. – Resultados de la predicción del intervalo 1, prueba 2.

4.3.3 Descripción de la predicción final dada en el intervalo 3

Una vez realizada la fase de calibración y entrenamiento la herramienta dará el resultado del modelo de predicción final, el cual fue indicado que se generará en el intervalo 3, con profundidad de 861 – 1160 m. La predicción dada fue mediante el uso del $C = 1000$, y $=.0001$ y se obtuvo un acierto de 98.66%, comparado con los datos reales que se le dieron.

Como se observa en la figura 20, las facies predominantes en los datos reales y predichos son de 3 (verde) y 4 (azul). La facies 4 (D), nos indica que la litología descrita es dada por un intercalación de rocas dolomita con alta concentración de vórgulos, concentración baja de fracturas y baja concentración de porosidad de matriz sin impregnación de lutita.

- En el intervalo 980 – 990 m, se observa presencia de la facies, que nos indica roca dolomita con caliza, media concentración de porosidad secundaria vugular, media concentración de porosidad por fracturas y alta concentración de porosidad primaria , donde existe impregnación de lutita.
- 1030 -1040 m, está comprendida por la facies 1 que indica presencia de roca caliza con lutita, fracturada y con porosidad primaria alta .
- 1100-1200 m, está comprendida, por caliza dolomitizada, con porosidad vugular sin presencia de lutita ni porosidad primaria.

La discretización de la porosidad proporcionada por los modelos de predicción es congruente con la descripción de facies ante descrita, ya que se obtienen porosidades de fractura, de vórgulos comunicados y de matriz (compacta), la roca predominante es dolomita pura, y dolomita con caliza.

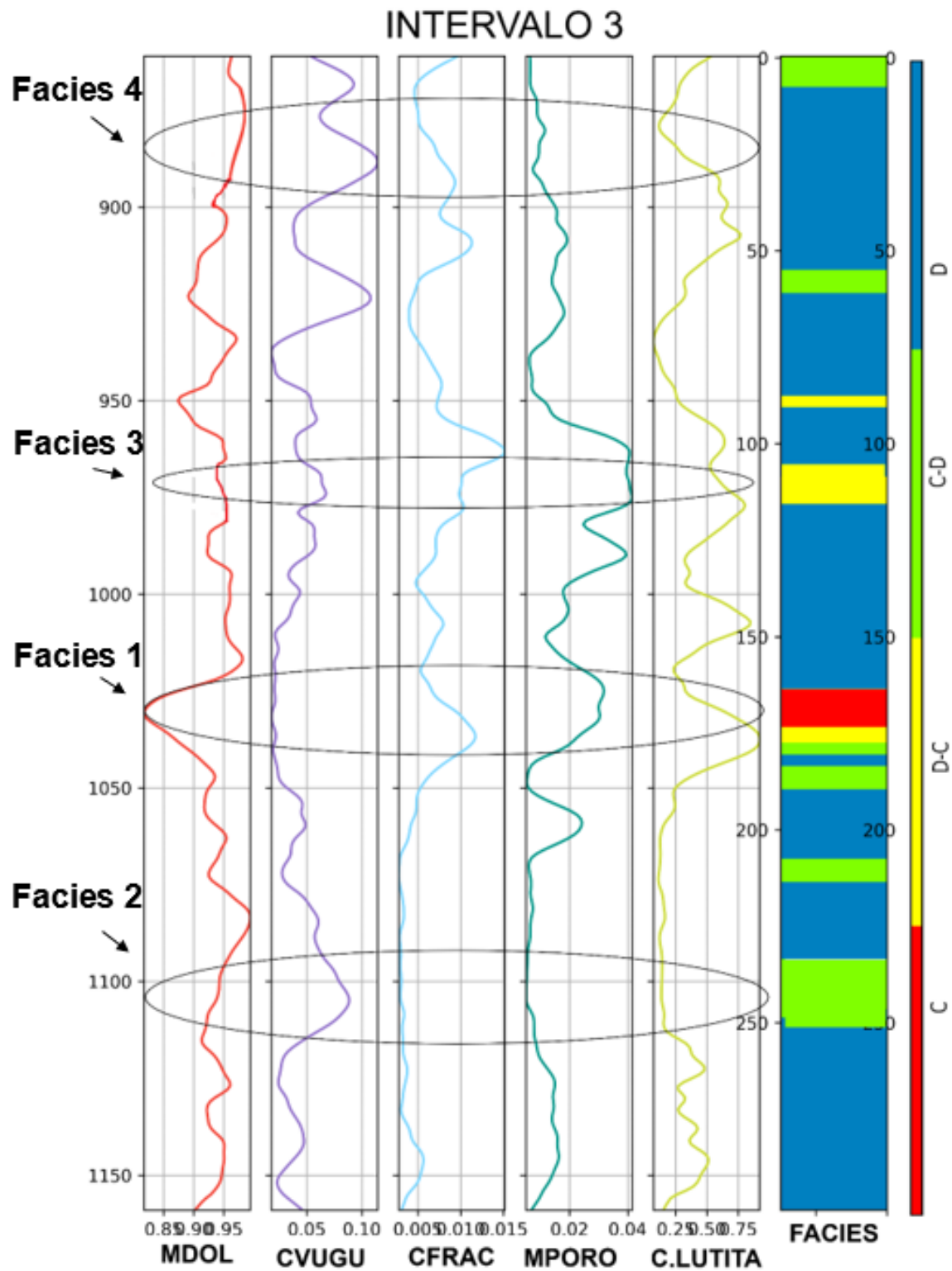


Figura 20. – Resultados de la predicción del intervalo 3, prueba 2.

4.4 Comparación de resultados de la prueba 1 y 2

La partición de porosidad que se obtiene con la herramienta IMP-Predict, es muy heterogénea, que va desde porosidades asociadas a matriz (compacta y permeable), vórgulos comunicados y aislados, así como microfracturas y fracturas. La técnica del K-means, proporciona resultados concordantes con las descripciones de las muestras de los registros, observándose zonas con porosidad secundaria está asociada a vórgulos aislados. Generalmente, las rocas carbonatadas con porosidad total menor al 7% tienen altas probabilidades de fracturarse, y en esta zona por medio de la aplicación de las técnicas se observa la presencia de fracturas, la técnica de asignación de porcentajes genera facies que indican la presencia de porosidad de matriz, vugular y fracturas en todo el intervalo, resaltando una similitud entre ambas técnicas de generación de facies, a pesar de que se tienen muy pocas diferencias tales como, pequeñas zonas de porosidad de vórgulos no conectados y presencia de lutita. de manera general la partición de la porosidad es congruente con las descripciones de las variables petrofísicas y con la información geológica que se dio en un previo análisis con el Instituto mexicano del petróleo.

En conclusión, realizando un análisis de ambas prueba el yacimiento carbonatado A, está conformada por brechas sedimentarias de dolomías (micro a meso cristalinas) con buena porosidad vugular (principalmente), con presencia fracturas abiertas. En ciertos intervalos están compuesta por dolomías con caliza con buena porosidad de matriz y cavidades de disolución, además en estas zonas existe un intervalo que resultó con alta concentraciones de lutitas. Las facies 1 corresponden a una caliza compacta, con porosidad de matriz y de fracturas (algunas cementadas con calcita). Las facies 3 y 4, se componen de dolomitas compactas, recristalizadas, con bajas porosidades de matriz y de microfracturas (algunas cementadas con calcita) y un bajo volumen de lutita.

4.5 Modelos de predicción Prueba 3 : yacimiento carbonatado B

Los modelos de predicción creados en esta prueba, fueron realizados con el objetivo de mostrar la efectividad, de la herramienta IMP-Predict, debido a que estos datos contienen facies reales que tienen como objetivo mostrar el tipo de roca predominante en el yacimiento, por motivos de discreción de datos propios del IMP, no se hace mención del nombre del pozo, de su ubicación y geología, por el cual se nombra a este yacimiento como prueba 3 o B.

4.5.1 Descripción de la predicción dada en el intervalo 1 y 3.

El yacimiento B, cuenta con un set de registros cuya calidad de información es buena, mostrando su comportamiento de cada curva al cambio de una facies litológica. La base de datos cuenta con una profundidad que va desde 2998.7 – 3404.5 m, donde existe un concentración de roca caliza, caliza con dolomita, dolomita calcárea y poca presencia de dolomita pura.

El comportamiento que se observa en las segunda predicción es la misma, de manera general se observa que cuando la curva de registro de rayos gamma llega a valores altos indica presencia de dolomita, es debido al mineral presente que dispara dicho registro, se observa que el registro resistivo (RT), decrece ante la presencia de dolomita y va aumentando al convertirse en roca caliza, es representa una presencia de porosidad en la roca.

En el intervalo 3080-3090 a 3220-3230, concentraciones alta de la facies 1, donde la respuesta del registro NPHI es bajo por la densidad de la calcita que es menos densa que la dolomita, debido a su contenido mineral,. Por otro lado el registro de tiempo de tránsito es bajo en presencia de roca caliza y alto en presencia de roca dolomita, ya que la roca dolomita es más compacta que la caliza.

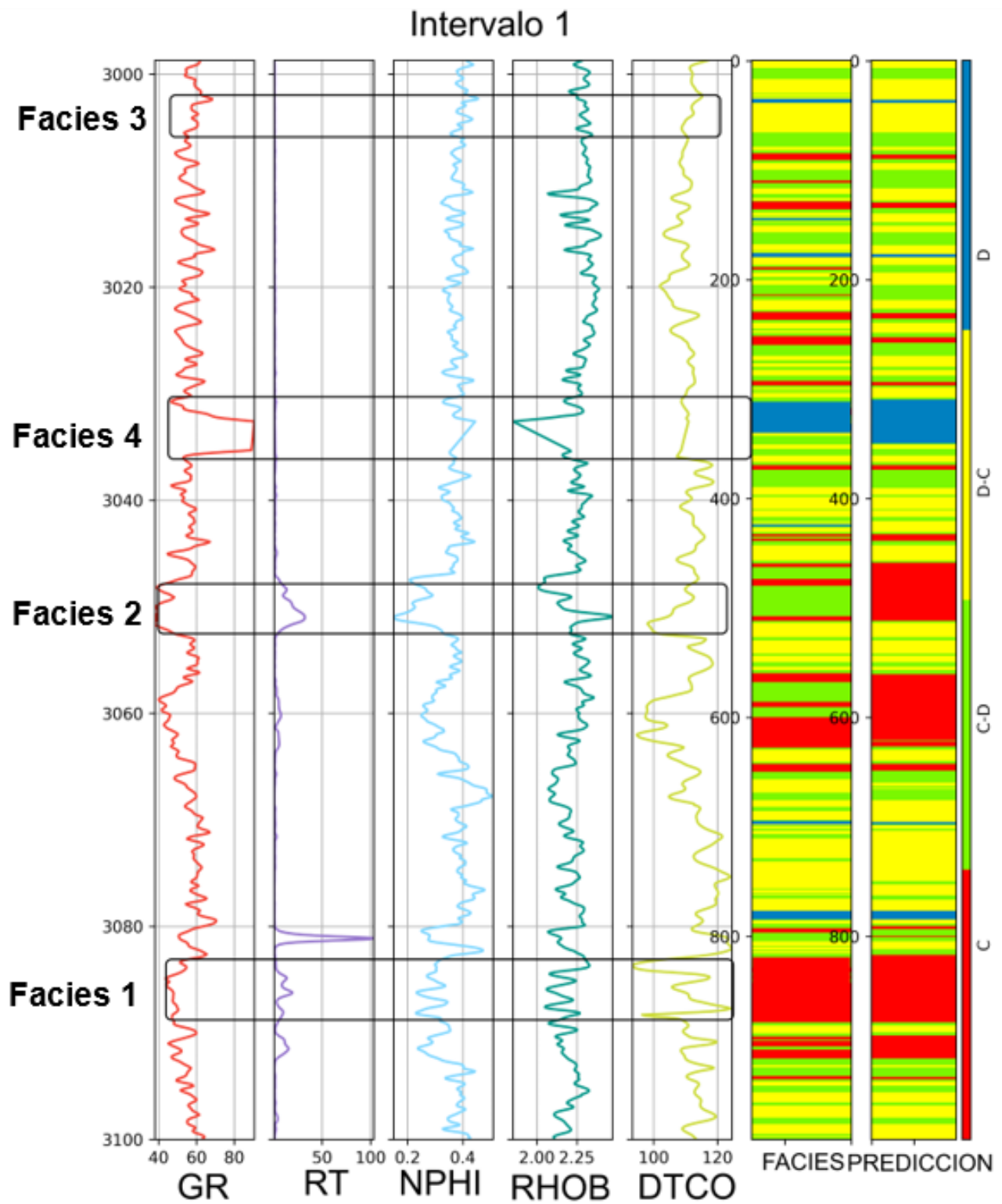


Figura 21. – Resultados de la predicción del intervalo 1 yacimiento B.

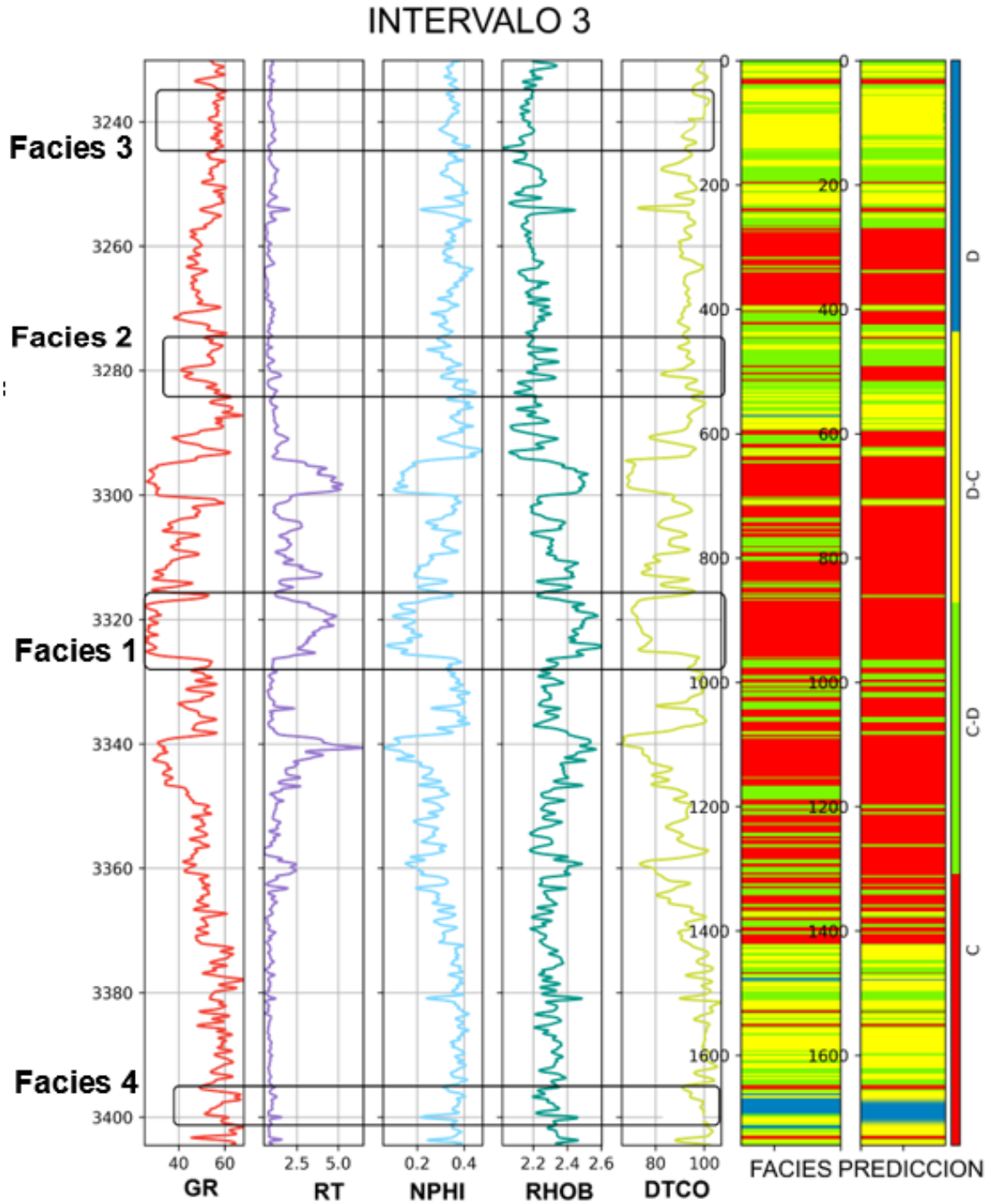


Figura 22. – Resultados de la predicción del intervalo 2, yacimiento B - prueba 3.

Para dar veracidad de la herramienta IMP-Predict, se hace esta evaluación de los modelos de predicción con las facies reales, lo cual me otorga un acierto del 85%, nos indica que la herramienta esta lista para usarse en casos reales en los cuales no se tenga información de un intervalo para poder predecirlo. En la parte de anexo se observaran los porcentajes dado por todos los C y y probados.

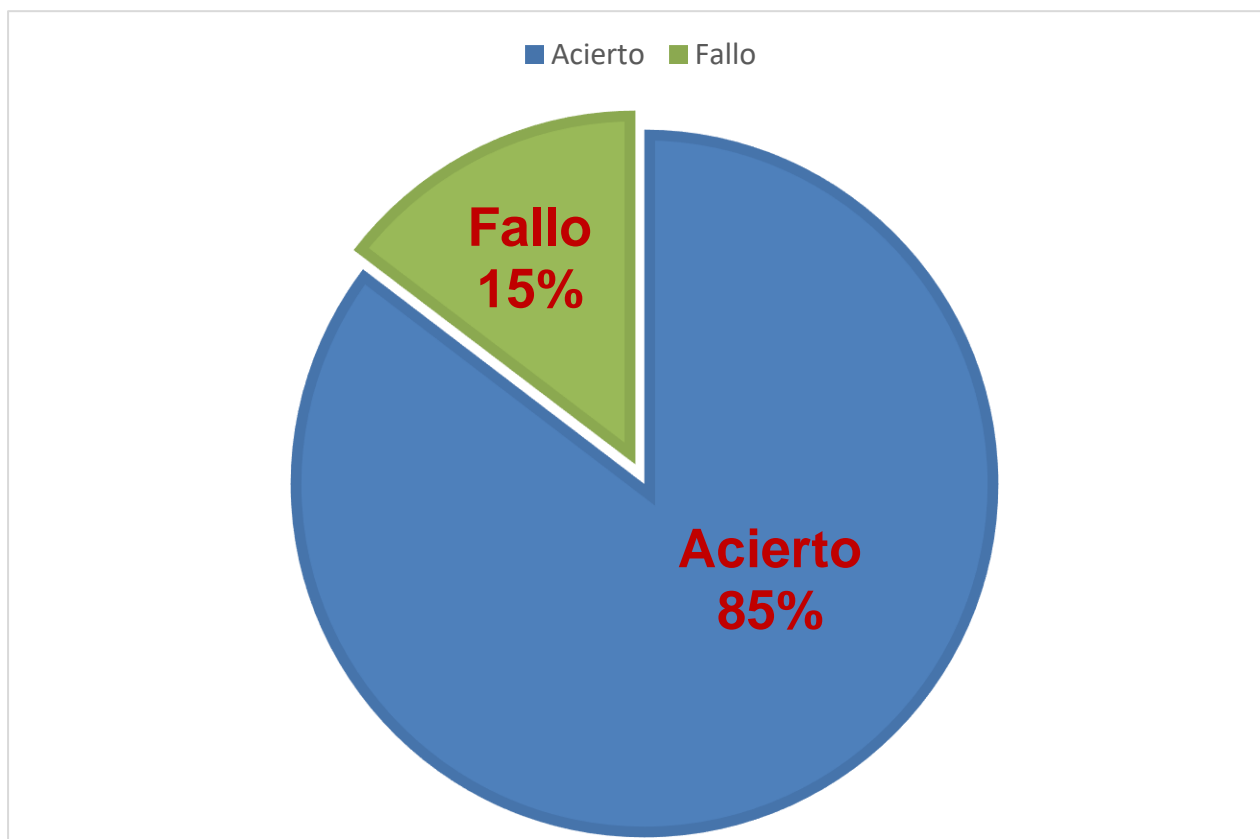


Figura 23. – Acierto general de la predicción de la prueba 3.

Capítulo 5. Conclusiones y recomendaciones

El presente trabajo abordó el interés que se ha dado en los últimos años al uso de la inteligencia artificial en el ámbito de la geofísica, con el análisis de dos casos de estudio. En ambos casos se buscó la metodología apropiada con el cálculo de facies y así obtener los modelos de predicción otorgados por la herramienta IMP-Predict V 2.0.

Para el desarrollo de los modelos de predicción de este tipo de yacimientos carbonatados se necesita más información y técnicas que nos ayuden a dar un mejor análisis geológico. Por otro lado es indispensable, aun siguiendo la hipótesis propuesta, el estudio y la experimentación, un análisis petrofísico para así obtener un mejor resultado en el comportamiento de estas rocas, sin embargo como se menciona solo son datos sintéticos utilizados para la comprobación de metodologías del cálculo de facies y el uso de la inteligencia artificial. El hecho de tomar información publicada por otros autores, con lleva a una incertidumbre, pues se desconocen algunos parámetros específicos y claves que nos indiquen como se obtuvieron esas facies, es por ello que en esta investigación se dan las siguientes conclusiones.

- La creación de un código que nos ayude a clasificar facies de manera automática es de gran ayuda para realizar correlación geológica con la zona de estudio y así poder definir una clasificación de facies.
- En los datos carbonatados del yacimiento A se encontraron 4 tipos de facies donde las rocas principales son: calizas, calizas con dolomitas, dolomitas calcáreas y dolomitas, mismas que fueron confirmadas por el IMP. Estos grupos de rocas presentan un sistema de triple porosidad comprendidas por porosidad primaria de matriz de roca, porosidades secundarias de fracturas y vóculos.
- Se comprueba la hipótesis acerca de obtener un mejor modelo de predicción utilizando la curva azul de Cross – Validation en los parámetros C Y Gamma dados por la herramienta.

- Los modelos de predicción propuestos en la tesis representan las facies en los yacimientos de carbonatos desde el punto de vista de la inteligencia artificial.
- Conocer la distribución de los tipos de porosidad (matriz, vórgulos y fracturas) que existen en el yacimiento, ayuda a la identificación de zonas con mejores características de explotación, ratifica intervalos de interés o propuestos y apoya a la selección de intervalos en subsecuentes pozos.
- De manera general la comprobación de los modelos de predicción propuestos utilizando datos sintéticos, demostraron la eficiencia de la herramienta para un caracterización de litologías por medio de facies, lo cual ayuda a dar una confiabilidad del uso de la herramienta para datos reales.
- Se recomienda al usuario, utilizar el criterio de un geólogo si se utiliza la herramienta IMP-Predict o el código de clasificación de facies no supervisado en datos reales, para así tener una comprobación optima de la información obtenida.
- Para reducir los niveles de incertidumbre en la caracterización de los sistemas porosos, se debe integrar información relacionada con núcleos, petrografía, diagénesis, ambientes de depósito e incluso registros de micro imágenes resistivas
- Finalmente en este trabajo se observó el potencial que tiene el software IMP-Predict v 2.0, aunque es una versión beta, se propone la lectura de más variables y la predicción no solo de facies litológicas, sino de facies sísmicas o la predicción de registros donde no se tiene información. De igual manera se debería implementar el uso de técnicas no supervisadas, lo cual ayudaría a no emplear muestra de aprendizaje y así se utilizarían cuando no existe información “a priori” sobre los objetos geológicos, o cuando esta información es muy reducida.

Bibliografías

- Aguilera, R. (1995). Naturally Fractured Reservoirs (Segunda ed.). Tulsa, Oklahoma, E.U.A.: PennWell Publishing Company.
- Arroyo, C.A. (1996) Base teóricas e Interpretación de Registros Geofísicos de Pozo. Universidad Nacional Autónoma de México. CDMX
- Archie GE: "Introduction to Petrophysics of Reservoir Rocks," Boletín de la AAPG 34, no. 5 (Mayo de 1950):943–961.
- Ávila, A. & Hernandez, O. (2012). Métodos para la estimación de electrofacies usando registros geofísicos de pozos convencionales (tesis de licenciatura). Instituto Politécnico Nacional. CDMX
- Bächle, G.T. & Harmersbach, Z.,(2009). Effects of pore structure on velocity and permeability in carbonate rocks. Germany: Tübingen University.
- BIJU-DUVAL, B. (2002). Sedimentary Geology "Sedimentary Basin, Depositional Environments, Petroleum Formation". París, Francia: TECHNIP
- Bratton, T., Canh, D. V., Que, N. V., Duc, N. V., Gillespie, P., Hunt, D., . . . Sonneland, L. (2006). "La naturaleza de los yacimientos naturalmente fracturados". Oilfield Review, 4-25.
- Brown, L.F.Jr. & W.L. Fisher.(1980). *Seismic stratigraphy interpretation and petroleum exploration*. Amer. Assoc. Petrol. Geol. Contin. Ed. Course Note Ser. 16, 125 p.
- Dubois, M. K., G. C. Bohling, & S. Chakrabarti, 2007, Comparison of four approaches to a rock facies classification problem: Computers & Geosciences, 33, no. 5, 599–617, <http://dx.doi.org/10.1016/j.cageo.2006.08.011>.

- Folk, R.L. Petrology of sedimentary rocks. Hemphillis, Austin. Texa. 1974. 170p.
- Geotem.(s.f).*Registros Geofísicos de Pozos*. CDMX. Recuperado de: <http://www.geotem.com.mx/registro.php>
- Gómez R. M & Ramírez H. R.,(2004). *Algoritmo para determinar velocidades a partir de porosidades en arenas, usando módulos elásticos y registros geofísicos de pozo*. (tesis de licenciatura). Instituto Politécnico Nacional. CDMX
- Hall, B. (2016). Facies Classification Using Machine Learning. *The Leading Edge*, Vol. 35, No. 10, Pg. 906 – 909.
- Humberto. S. (2012). Evaluación de la calidad del aislamiento de alto voltaje utilizando máquinas de soporte vectorial (tesis de maestría). CIATEC. León, Guanajuato
- Kamath, N., Tsvankin, I., Naeini, E. Z. (2017). Facies – Constrained FWI: Toward Application to Reservoir Characterization. *The Leading Edge*, Vol. 36, No. 11, Pg. 924 – 930.
- Lucia, J. F. (2007). "Carbonate Reservoir Characterization (Segunda ed.). Austin, Texas, Estados Unidos de América: Springer.
- Martínez, A. (2005). Conceptos sedimentológicos, estratigráficos y estructurales en la geología de yacimientos de fluidos (tesis de licenciatura). Universidad Nacional Autónoma de México. CDMX
- Python. (En línea). Disponible en: <https://www.python.org>
- Rodrigo, J.A.(2017). Máquinas de Vector Soporte (Support Vector Machines, SVMs). RPUBS POR R STUDIO. https://rpubs.com/Joaquin_AR/267926

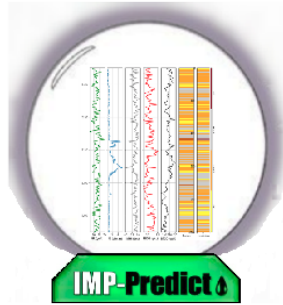
- Romero, M., Díaz, M., Del Valle, R., Albino, J. (2018). Guía De Usuario de la Herramienta Computacional IMP – Predict Versión 2.0. Instituto Mexicano del Petróleo.
- Sayaka, P.(2018). K-Means Clustering en Python con scikit-learn. DATA CAMP. <https://www.datacamp.com/community/tutorials/k-means-clustering-python>
- Serra, O.& H. T. Abott,(1982). *The contributions of loggings data to sedimentology and stratigraphy*: Society of Petroleum Engineers Journal, 22, stratigraphy, pp 117 131.
- Scikit Learn. (En línea). Disponible en: <http://scikit-learn.org>
- Schlumberger. (2008). *Principios/Aplicaciones de la interpretación de Registros*.
- Tarbuck, E.J. & Lutengs, F.K.,y Tasa D.(2005).*Ciencias de la tierra: una introducción a la geología física*.(8° ed.). Madrid: Pearson Educación S.A.
- Tiab, D., & Donaldson, E. (2004). Petrophysics." Theory and Practice of Measuring Reservoir Rock and Fluid Transport Properties", (Segunda ed.). E.U.A.: Gulf Professional Publishing.

Anexos

Tutorial del Procesamiento en el software IMP-Predict V 2.0

En el siguiente apartado se mostraran los pasos a seguir para el procesamiento de los datos tratados anteriormente, con el fin de dar al lector una breve guía de cómo es el entorno del software, por consiguiente los pasos son:

1. Iniciar la herramienta IMP-Predict para lo cual se deberá ubicar el archivo IMP Predict_V2.0.exe y ejecutarlo. Una vez lanzada la ejecución, se mostrará la pantalla de bienvenida (ver Figura 24), después se muestra la pantalla principal de la herramienta (ver Figura 25) que indica el punto de partida del procesamiento.
2. En su pantalla principal se muestran 4 menús. Se seleccionará el menú archivo y dentro del menú Archivo existen cuatro opciones relacionadas con el proyecto de predicción como se muestra en la Figura 25. Se seleccionará la opción de nuevo proyecto predicción donde se desplegará una ventana (ver Figura 26), a través de la cual el usuario podrá navegar por su sistema de archivos con el fin de seleccionar los archivos de entrenamiento y de predicción que formarán parte del proyecto para la construcción del modelo de predicción. Asimismo, podrá seleccionar una ruta en donde se guardarán todos los resultados generados. se podrá seleccionar la opción si desea comparar la predicción resultante con datos reales. Dichos datos deberán contener todas las facies reales correspondientes al archivo predicción. Por último, se desplegará una opción en donde se pregunta al usuario si desea visualizar los resultados al finalizar el procesamiento.



IMP-Predict

PREDICCIÓN DE FACIES A ESCALA DE
REGISTROS DE POZO

Versión 2.0

Gerencia de Geofísica Cuantitativa
Instituto Mexicano del Petróleo

Septiembre 2017

Herramienta computacional desarrollada en el Proyecto Proyecto D.61037
“Interpretación Sísmica Cuantitativa Guiada por Facies Litológicas para la
Caracterización de Yacimientos” del Comité de Investigación, Innovación y Soluciones
(CIIS) - Copyright (c) 2017 Instituto Mexicano del Petróleo.

Figura 24. - Pantalla de bienvenida de la herramienta computacional IMP-Predict Versión 2.0

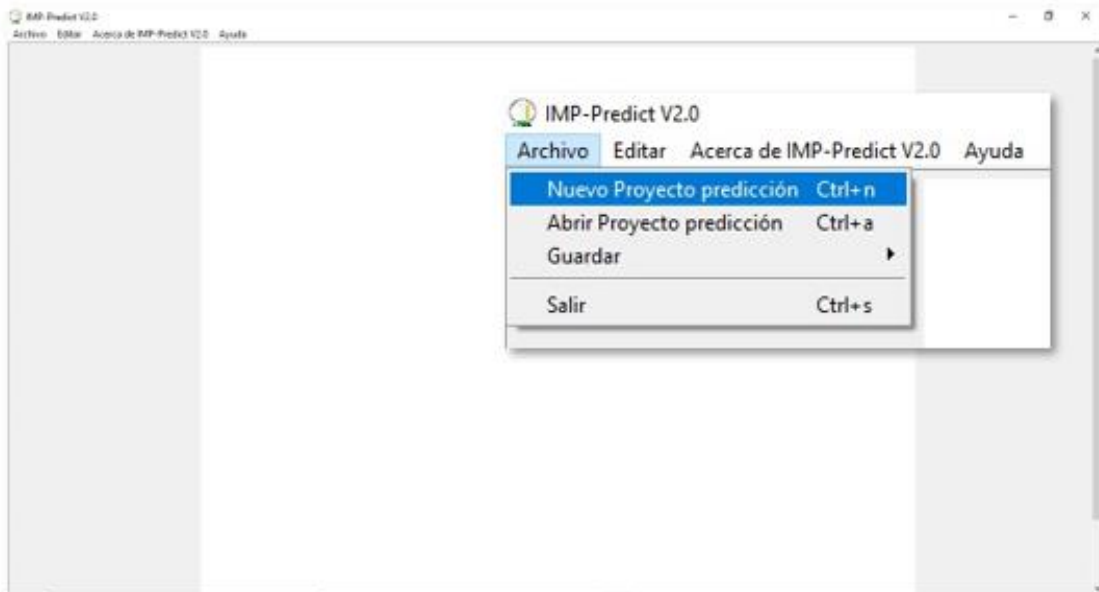


Figura 25. - Pantalla principal de la herramienta computacional IMP-Predict Versión 2.0

- Una vez seleccionados los archivos indicados, se procede a continuar con la configuración del proyecto oprimiendo el botón correspondiente o también se puede cancelar la edición y volver a la pantalla principal. Se procederá a abrir una ventana de selección de parámetros (ver Figura 27) en donde se podrá seleccionar el pozo o intervalo de pozo que funcionará como oculto para la calibración del modelo de predicción y los pozos o intervalos de pozo que se utilizarán para el entrenamiento del modelo, asimismo se mostrará la lista de pozos o intervalos de pozos susceptibles a predecir donde se visualizará la lista de variables predictoras que serán utilizadas para la construcción del modelo de predicción.

Edición del Proyecto predicción

Seleccionar el archivo de datos de entrenamiento

Seleccionar el archivo de datos a aplicar la predicción

Seleccionar la ruta para guardar los resultados

¿Desea visualizar una comparación con datos reales?

Marque la casilla si cuenta con los datos reales

¿Desea visualizar la bitácora de resultados?

Marque la casilla si desea verla al terminar procesamiento

Figura 26. - Ventana de edición del proyecto predicción.

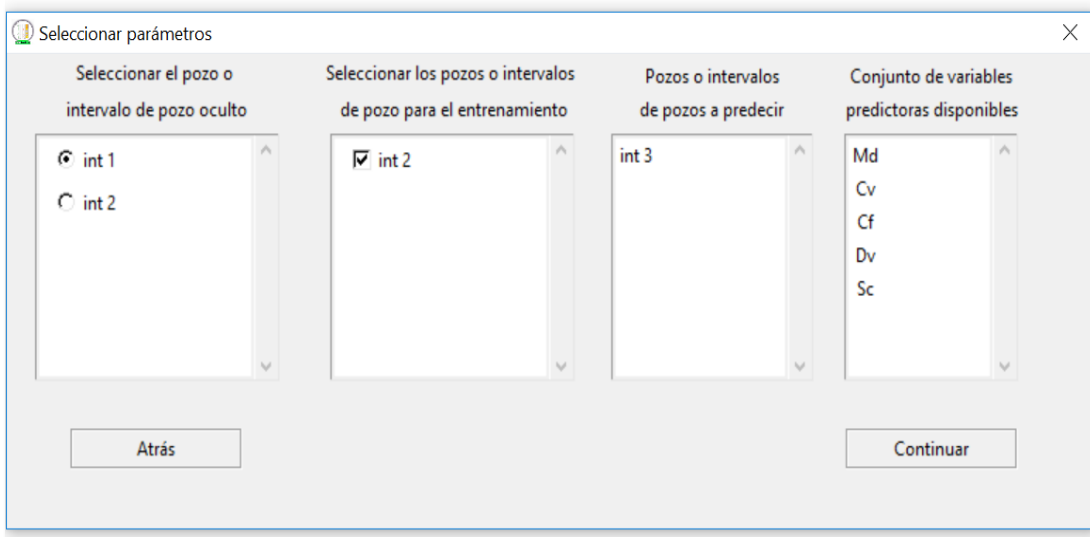


Figura 27. - Ventana de selección de parámetros.

- Una vez realizada la selección de los distintos parámetros, se procede a continuar con la configuración del proyecto oprimiendo el botón correspondiente o también es posible volver a la ventana anterior oprimiendo el botón atrás. Enseguida, se mostrará la ventana en donde se deberán seleccionar el(los) pozo(s) o intervalo(s) de pozo que se va(n) a predecir (ver Figura 28), así como la clase (parámetro a predecir), es decir las facies.

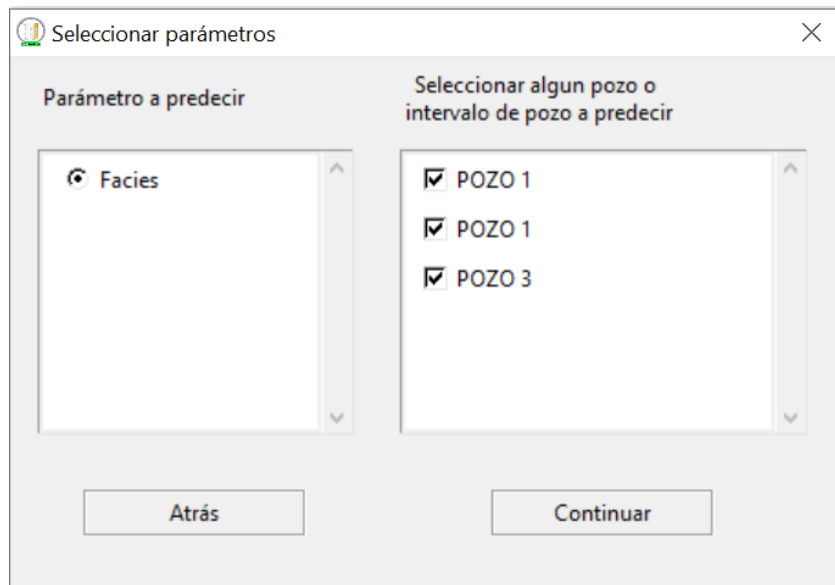


Figura 28. - Ventana de selección de parámetros a predecir.

5. Una vez realizada la selección de los parámetros de predicción, se procede a continuar con la configuración del proyecto oprimiendo el botón continuar o también es posible volver a la ventana anterior oprimiendo el botón Atrás. Enseguida, se presentará la ventana (ver Figura 29) que muestra las facies adyacentes presentes en todos los datos de entrenamiento y su proporción. En esta misma ventana, se mostrará un apartado que muestra la lista de facies adyacentes encontradas y permite seleccionar aquellas que se deseen eliminar con el fin de que no sean tomadas en cuenta para la construcción del modelo de predicción. Asimismo, existe la posibilidad de poder volver a la lista original usando el botón Restaurar.

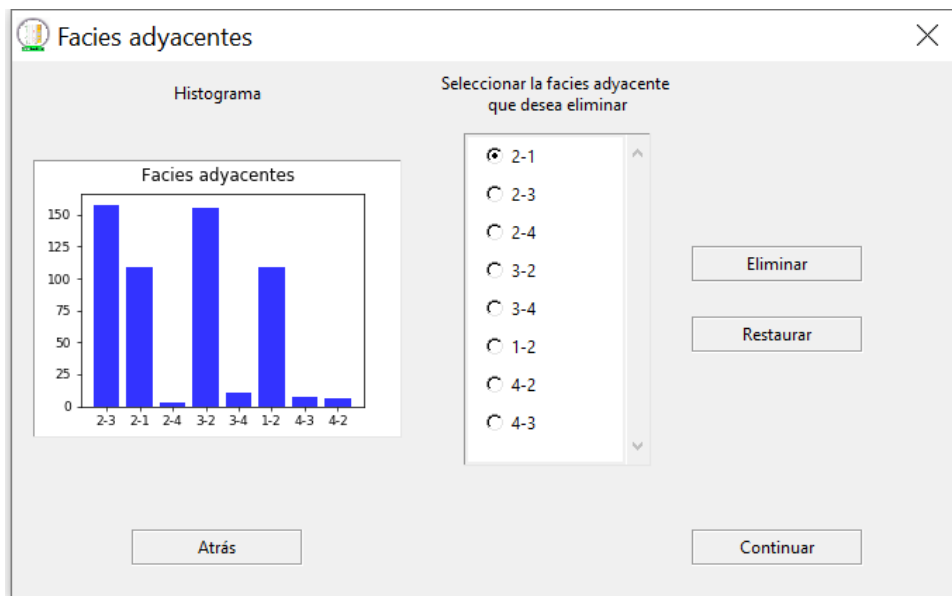


Figura 29. - Ventana de facies adyacentes.

6. Una vez revisada y actualizada la lista de facies adyacentes, se continua con la configuración del proyecto oprimiendo el botón correspondiente o también es posible volver a la ventana anterior oprimiendo el botón Atrás. Se presentará la ventana de personalización de los colores que serán utilizados para las gráficas y los nemónicos que identificarán a las variables predictoras como se muestra en la Figura 30.

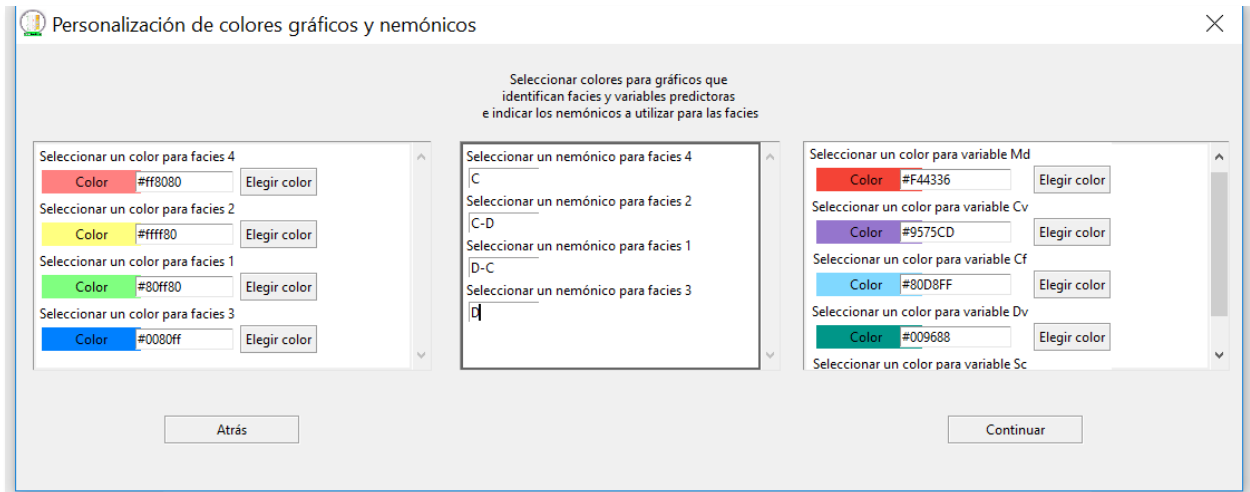


Figura 30. - Ventana de personalización de colores para gráficos y nemónicos de facies.

7. Una vez seleccionados los colores para gráficos y nemónicos de facies, se procede a continuar con la configuración del proyecto, donde aparecerá una ventana, esta indica el inicio de la construcción del modelo de predicción, como se muestra en la Figura 31. Una vez transcurrido el tiempo de procesamiento, se presentará una ventana preguntando si se desean visualizar las gráficas de C y Gamma. Si el usuario selecciona la opción de visualizar las gráficas, se desplegarán seis gráficas con diferentes valores de Gamma y posteriormente, aparecerá una ventana en donde el usuario podrá ingresar los valores de C y Gamma que considere más adecuados para la construcción del modelo de predicción (ver en la Figura 32).

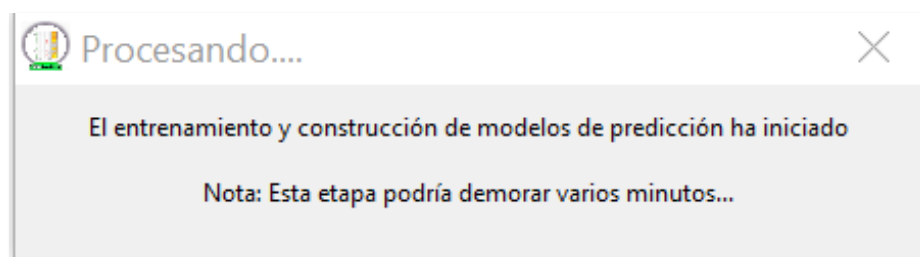


Figura 31. - Ventana de aviso de inicio del procesamiento.

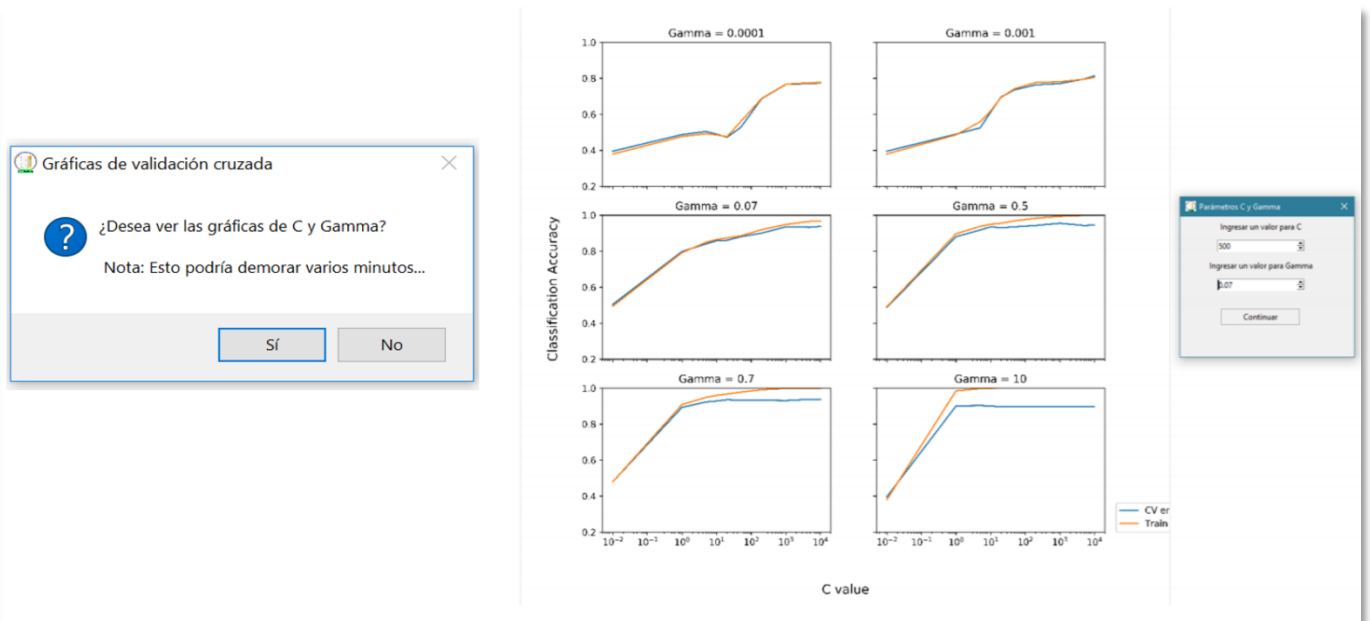


Figura 32. - Ventanas de visualización de gráficas de C , Gamma y de ingreso de parámetros.

- En caso de que se haya seleccionado no visualizar las gráficas C y Gamma, se mostrará únicamente la ventana que permite ingresar los parámetros C y Gamma por default, o de haber ingresado los valores de C y Gamma, se procede a continuar con el procesamiento, oprimiendo el botón correspondiente, lo cual permitirá aplicar el modelo de predicción al conjunto de datos a predecir. De esta manera, aparecerá una ventana de aviso que indica que el procesamiento continúa (ver Figura 33).

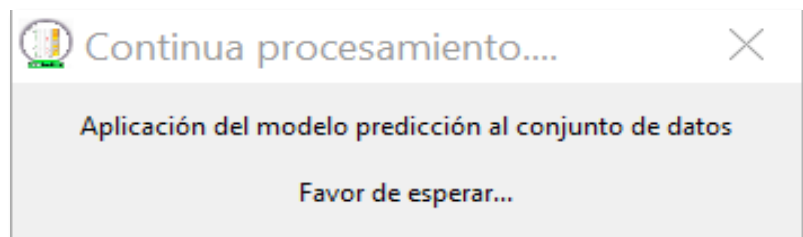


Figura 33. - Ventana de aviso continuación del procesamiento

- Como último paso, se presentarán los resultados generados en la pantalla y si en la ventana de Edición del proyecto predicción se marcó la casilla para desplegar la bitácora, al finalizar el procesamiento, se mostrará un resultado como se muestra

en la Figura 34. Enseguida aparecerá una ventana (ver Figura 35) preguntando si se desea generar el documento PDF que contendrá el reporte de todas las gráficas generadas como resultado del proyecto de predicción. En caso de no haber marcado la casilla para desplegar la bitácora o registro de resultados, se visualizarán solamente los resultados del proyecto de predicción, así como la ventana de creación del documento PDF.

```

Log: Bloc de notas
Archivo Edición Formato Ver Ayuda

Start time: 00:00:00

IMP-PREDICT, Versión 2.0
Gerencia de Geofísica Cuantitativa
Copyright (c) 2017 Instituto Mexicano del Petróleo

Este código se ejecuta usando:
- Python version 3.6.0 |Continuum Analytics, Inc.| (default, Dec 23 2016, 11:57:41) [MSC v.1900 64 bit (AMD64)]
- Pandas version 0.19.2
- Matplotlib version 2.0.0

-----Tabla de datos leídos (extracto):-----
  Facies      Formation Well Name  Depth      Md      Cv      Cf \
0      4 Dato_formacion    int 1      1  0.60969  0.144530  0.005240
1      4 Dato_formacion    int 1      2  0.61846  0.146010  0.005341
2      4 Dato_formacion    int 1      3  0.62685  0.147620  0.005451
3      4 Dato_formacion    int 1      4  0.63415  0.149450  0.005554
4      4 Dato_formacion    int 1      5  0.64037  0.151440  0.005643
..      ...
855     2 Dato_formacion    int 2    856  0.95966  0.046977  0.011181
856     2 Dato_formacion    int 2    857  0.96351  0.044414  0.011337
857     2 Dato_formacion    int 2    858  0.96595  0.043461  0.011309
  
```

Figura 34. - Despliegue de bitácora o registro de resultados.

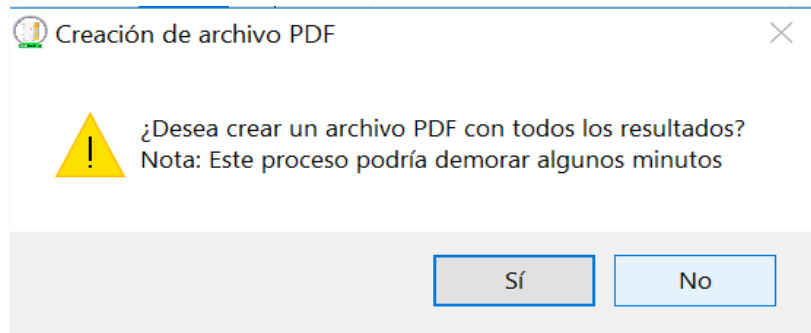


Figura 35. - Ventana de confirmación de creación del reporte de resultados en PDF.

10. Si el usuario selecciona crear el reporte de resultados en formato PDF, se mostrará una pequeña ventana notificando la creación del mismo (ver Figura 36) y posteriormente se presenta una ventana de aviso sobre la finalización de creación del reporte PDF. Posteriormente, una vez creado el reporte de resultados en formato PDF, se oprime el botón Aceptar, lo cual dará paso a la apertura y visualización de dicho documento. Cuando todos los procesos anteriores hayan finalizado, el usuario podrá visualizar los resultados en la pantalla principal

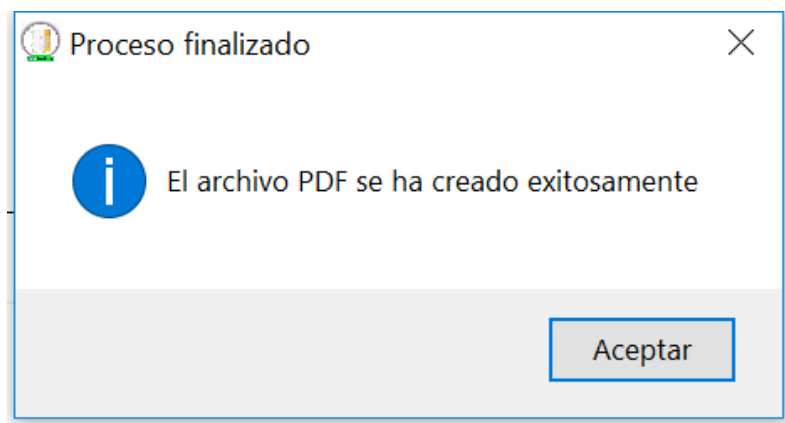


Figura 36. - Ventana de aviso de creación exitosa del reporte de resultados en formato PDF.

Código implementado para la clasificación por medio de K-means (Programa A)

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans

filename = 'Propiedades físicas'

# Reads the file to normalize
df = pd.read_csv(filename + '.csv')

# Exports the resultant dataframe
df.to_csv(filename + '_CLASSIFIED.csv', index=None, header=True)

columns = ['A.Matrix.Dolomite_C',
           'A.P1.Concentration',
           'A.P2.Concentration',
           'A.P1.Alpha1',
           'Shale.Concentration'
          ]

x = np.asanyarray(df[columns])

number_of_classes = 3

model = KMeans(init='k-means++', n_clusters=number_of_classes)

model.fit(x)

df['Facies'] = 0

for i in range(df['Depth'].count()):
    df.loc[i, 'Facies'] = model.labels_[i]+1

# Exports the resultant dataframe
df.to_csv(filename + '_CLASSIFIED.csv', index=None, header=True)

file = open('log.txt', 'w+')

file.write('*** Cantidades de registros de cada Facie ***')

for i in range(1,number_of_classes+1):
    count = df[df['Facies'] == i]['Facies'].count()
    file.write('\n' + str(i) + ': ' + str(count))
```

```

file.close()

df = df.iloc[0:0]

for i in range(number_of_classes):
    new_columns = columns + ['Facies']
    cluster_centers = model.cluster_centers_[i].tolist() + [i+1]
    df.loc[i, new_columns] = cluster_centers

df.to_csv(filename + '_CLUSTER_CENTERS.csv', index=None, header=True)

```

Código para la clasificación de facies por medio de porcentajes (programa B)

```

import pandas as pd

def verify_properties(headers, keys):
    for key in keys:
        if key not in headers:
            raise Exception("{} doesn't exist in the headers".format(key))

def verify_properties_ranges(dictionary):
    for key, value in dictionary.items():
        if len(value) < 2:
            raise Exception("{} has less than 2 ranges".format(key))

def sort_properties_ranges(dictionary):
    for key in dictionary.keys():
        dictionary[key] = sorted(dictionary[key])

    return dictionary

def calculate_bits(dictionary):
    bits = 0
    for key, value in dictionary.items():
        two_power = 1
        while(len(value) - 1 > 2 * two_power):
            two_power += 1
        bits += two_power

    return bits

```

```

def get_bits_fixed(value, range):
    bits = bin(value)[2:]
    max_bits = get_max_bits(range)

    while(len(bits) < max_bits):
        bits = '0' + bits

    return bits

def get_facie(row, dictionary):
    bits_str = ""

    for key, value in dictionary.items():
        for i in range(len(dictionary[key]) - 1):
            if row[key] >= dictionary[key][i] and row[key] <= dictionary[key][i + 1]:
                bits_str += get_bits_fixed(i, dictionary[key])
                break

    return int(bits_str, 2)

def get_max_bits(range):
    two_power = 1

    while(len(range) - 1 > 2 * two_power):
        two_power += 1

    return two_power

def get_facies_description(facie, bits_length, dictionary):
    bits = bin(facie)[2:]
    while(len(bits) < bits_length):
        bits = '0' + bits
    description = ""
    for key, value in dictionary.items():
        description += key + ' '
        max_bits = get_max_bits(value)
        x = int(bits[:max_bits], 2)
        bits = bits[max_bits:]
        description += ' [{} , {}], '.format(value[x], value[x + 1])

    return description

def classify_file(filename, dictionary):

```

```

df = pd.read_csv(filename + '.csv')
verify_properties(df.columns.values.tolist(), dictionary.keys())

verify_properties_ranges(dictionary)

sort_properties_ranges(dictionary)

bits = calculate_bits(dictionary)

df['Facies'] = 0

for i in range(df['Depth'].count()):
    df.loc[i, 'Facies'] = get_facies(df.loc[i], dictionary)

file = open('log.txt', 'w+')

file.write('*** Cantidades de registros de cada Facies ***')

facies_aux = []
facies = []
facie = 1
for i in range(2 ** bits):
    count = df[df['Facies'] == i]['Facies'].count()
    if count != 0:
        facies_aux.append(i)
        facies.append(facie)
        file.write('\n' + str(facie) + ': ' + str(count))
        facie += 1

file.write('\n\n*** Descripción de cada Facies ***')

for i in range(len(facies_aux)):
    description = get_facies_description(facies_aux[i], bits, dictionary)
    file.write('\n' + str(facies[i]) + ': ' + description)

file.close()

for i in range(df['Depth'].count()):
    for j in range(len(facies_aux)):
        if facies_aux[j] == df.loc[i, 'Facies']:
            df.loc[i, 'Facies'] = facies[j]

```

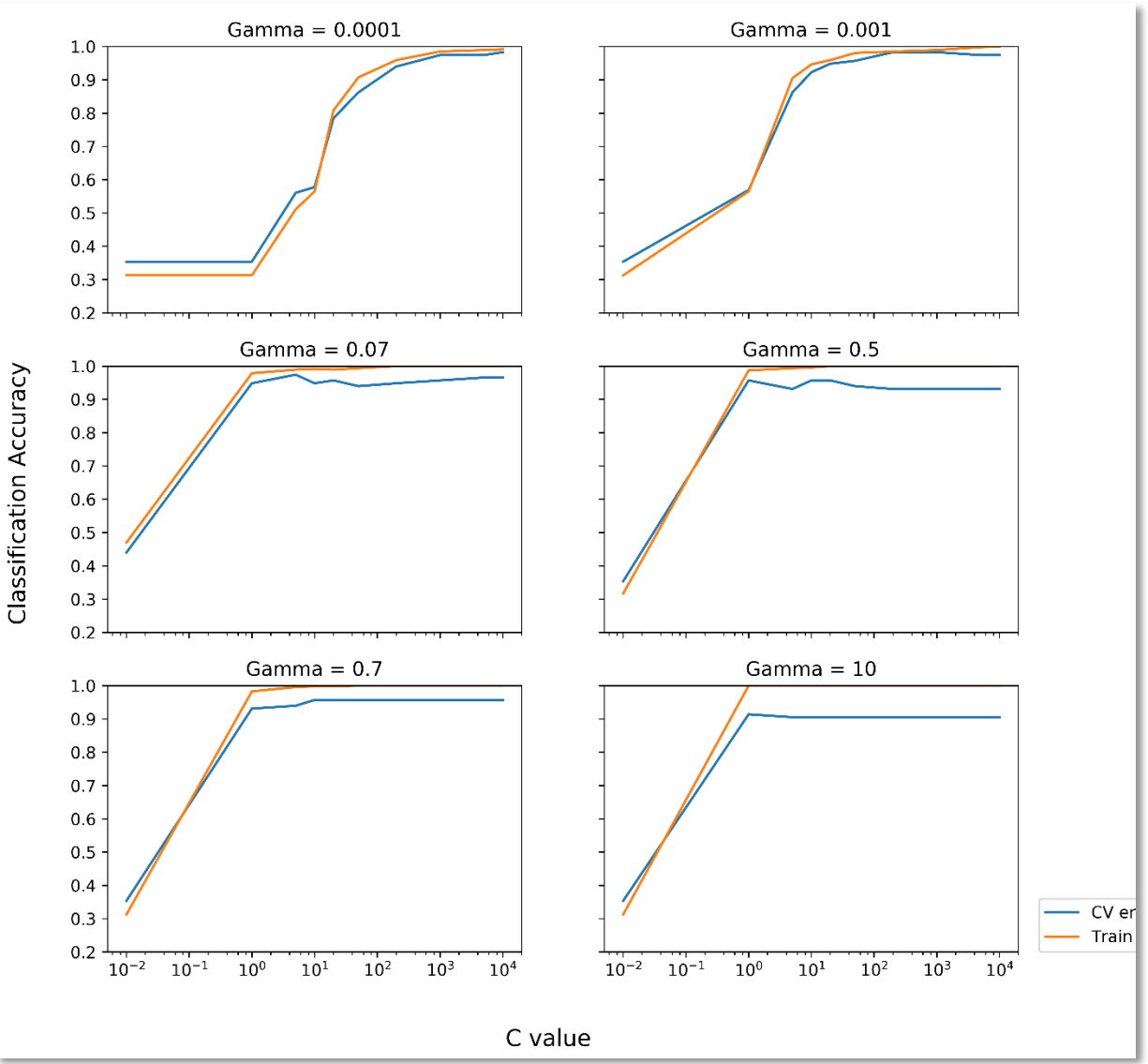


Figura 37. – Gráficos C Y GAMMA generados de la prueba 1.

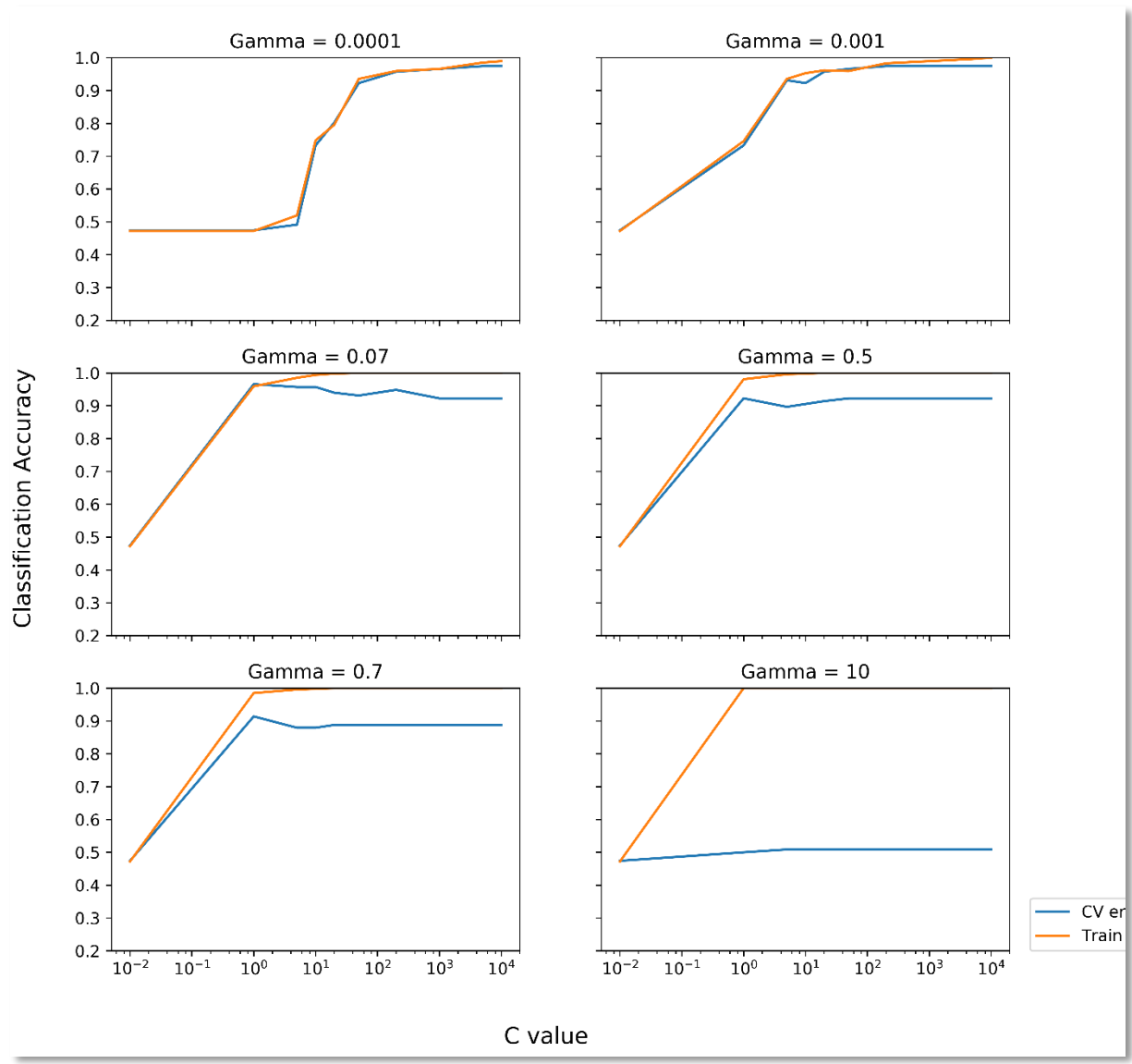


Figura 38. – Gráficos C Y GAMMA generados de la prueba 2.

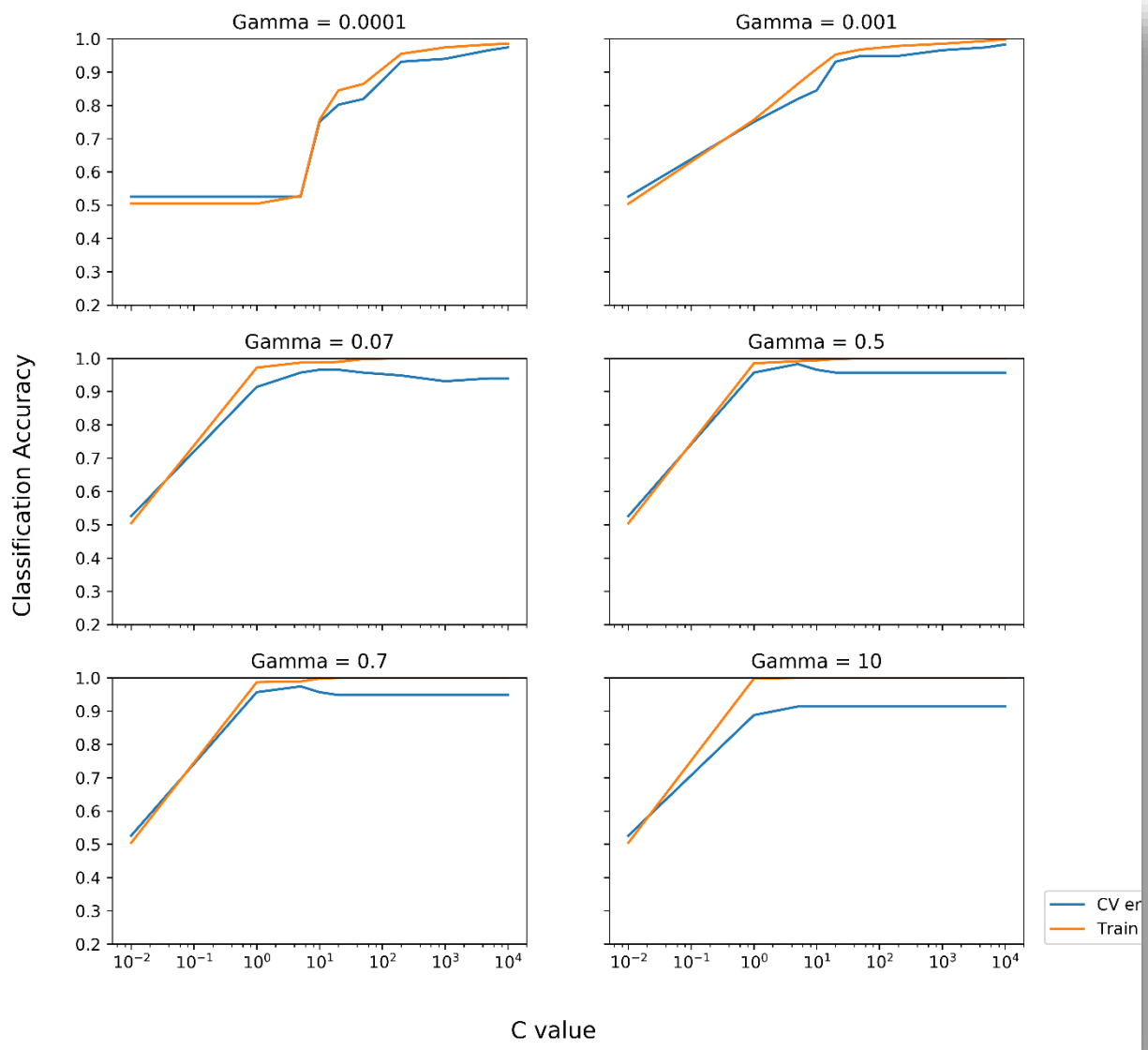


Figura 39. – Gráficos C Y GAMMA generados de la prueba 3.

C Y GAMMA	GENERAL
C:500 Y:0.07	85.36%
C:500 Y:0.0001	78.57%
C:500 Y:0.001	81.07%
C:500 Y:0.5	84.64%
C:500 Y :0.7	81.79%
C:500 Y:10	68.21%
Promedio	79.94%
C: 1000 Y :0.07	85.54%
C: 1000 Y :0.0001	79.29%
C: 1000 Y :0.001	80.71%
C: 1000 Y :0.7	81.07%
C: 1000 Y :0.5	82.50%
C: 1000 Y :10	68.21%
Promedio	79.55%

Tabla 14. Acierto de las 12 pruebas generadas de la prueba 3.