



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 1

Neiva, 23 de enero del 2020

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El (Los) suscrito(s):

Juan Daniel Vargas Perdomo, con C.C. No. 1075309707,

German Camilo Arce Calderón, con C.C. No. 1075301825.

Autor(es) de la tesis y/o trabajo de grado titulado “Determinación de la presión onset en yacimientos petrolíferos con redes neuronales implementadas en Python basado en modelo estadístico” presentado y aprobado en el año 2019 como requisito para optar al título de “Ingeniero de Petróleos”

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales “open access” y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE:

EL AUTOR/ESTUDIANTE:

Firma:

Firma:

Vigilada Mineducación



TÍTULO COMPLETO DEL TRABAJO:

AUTOR O AUTORES:

Primero y Segundo Apellido	Primero y Segundo Nombre
VARGAS PERDOMO	JUAN DANIEL
ARCE CALDERÓN	GERMAN CAMILO

DIRECTOR Y CODIRECTOR TESIS:

Primero y Segundo Apellido	Primero y Segundo Nombre
SEPÚLVEDA GAONA	JAIRO ANTONIO
MOSQUERA CERQUERA	VLADIMIR

ASESOR (ES):

Primero y Segundo Apellido	Primero y Segundo Nombre
----------------------------	--------------------------

PARA OPTAR AL TÍTULO DE: INGENIERO DE PETRÓLEOS

FACULTAD: INGENIERÍA

PROGRAMA O POSGRADO: INGENIERÍA DE PETRÓLEOS

CIUDAD: NEIVA

AÑO DE PRESENTACIÓN: 2019

NÚMERO DE PÁGINAS: 69

TIPO DE ILUSTRACIONES (Marcar con una X):

Diagramas___ Fotografías___ Grabaciones en discos___ Ilustraciones en general___ Grabados___
Láminas___ Litografías___ Mapas___ Música impresa___ Planos___ Retratos___ Sin ilustraciones X
Tablas o Cuadros___



SOFTWARE requerido y/o especializado para la lectura del documento:

MATERIAL ANEXO:

PREMIO O DISTINCIÓN (En caso de ser LAUREADAS o Meritoria):

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

<u>Español</u>	<u>Inglés</u>
1. Presión Onset	Onset Pressure
2. Red Neuronal	Neural network
3. Conexión Completa	Complete connection
4. Inteligencia Artificial	Artificial Intelligence
5. Python	Python
6. Temperatura	Temperatura
7. Presión de saturación	Saturation pressure
8. Máquinas de soporte vectorial	Support vectore machine

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

En este trabajo se desarrolló e implementó una red neuronal totalmente conectada (full connect) la cuál es llamada AsphPy, y un algoritmo de “Machine Learning” con la finalidad de predecir la presión Onset en los yacimientos de petróleos. Esta red neuronal fue elaborada en el lenguaje de programación Python y está basada en funciones estadísticas para el elaboración y evaluación del rendimiento del modelo de predicción. El algoritmo de Machine Learning implementó un soporte de maquina vectorial (SVM, Support Vector Machine), que tuvo como finalidad comparar sus resultados con el rendimiento de la red AsphPy.

Una finalidad importante de este trabajo es el uso de Python en modelos de predicción de la presión Onset realizados por el grupo de investigación COFA. El primer modelo,



DESCRIPCIÓN DE LA TESIS Y/O TRABAJOS DE GRADO

CÓDIGO	AP-BIB-FO-07	VERSIÓN	1	VIGENCIA	2014	PÁGINA	3 de 4
---------------	---------------------	----------------	----------	-----------------	-------------	---------------	---------------

presentado por Cristian Stiven Loaiza Cano y Samuel Quintero Herrera en el trabajo “Desarrollo de un Modelo para la Determinación de la Presión de Inicio de Precipitación de Asfáltenos (Onset) en Yacimientos de Petróleo”, es la base para la elaboración de AsphPy. El segundo modelo, presentado por Nathalie Flor Olave y Nicolas Hernando Quesada Bahamon en el trabajo “Determinación de la Presión Onset Usando Análisis Estadístico ANOVA en el Modelo de la Red Neuronal” desarrollaron mejoras al poder implementar la red neuronal basándose en un análisis ANOVA; el cual brindo información de gran ayuda para observar qué componentes o datos de entrada son más relevantes para la determinación de la presión Onset.

ABSTRACT: (Máximo 250 palabras)

In this work, a fully connected neural network (full connect), called AsphPy, and a machine learning algorithm were developed and implemented in order to predict Onset pressure in oil reservoirs. This neural network was developed in the Python programming language and is based on statistics for the evaluation of the performance of the artificial neural network. The Machine Learning algorithm implemented a support Vector Machine (SVM), which was intended to compare its results with the performance of the AsphPy network.

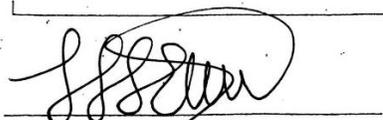
An important purpose of this work is the use of Python in Onset pressure prediction models by the COFA research group. The first model, presented by Cristian Stiven Loaiza Cano and Samuel Quintero Herrera in the work "development of a model for determining the asphaltene onset pressure in oil reservoirs", is the basis for the creation of AsphPy. In the second model, presented by Nathalie Flor Olave and Nicolas Hernando Quesada Bahamon in the work "determination of Onset pressure using ANOVA statistical analysis in neural network model", improvements were developed in the ability to implement the neural network based on an ANOVA analysis; which provided helpful information to see which components or input data are most relevant to the determination of Onset pressure.



CÓDIGO	AP-BIB-FO-07	VERSIÓN	1	VIGENCIA	2014	PÁGINA	4 de 4
--------	--------------	---------	---	----------	------	--------	--------

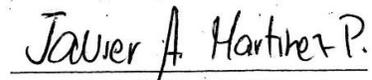
APROBACIÓN DE LA TESIS

Nombre Presidente Jurado: Freddy Humberto Escobar Macualo


FIRMA DEL JURADO

Firma:

Nombre Jurado: Javier Andres Martinez


FIRMA DEL JURADO

Firma:

**DETERMINACIÓN DE LA PRESIÓN ONSET EN YACIMIENTOS
PETROLÍFEROS CON REDES NEURONALES IMPLEMENTADAS EN
PYTHON BASADO EN MODELO ESTADÍSTICO.**



**JUAN DANIEL VARGAS PERDOMO
GERMÁN CAMILO ARCE CALDERÓN**

**UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
GRUPO DE INVESTIGACIÓN COMPORTAMIENTO DE FASES
PROGRAMA DE INGENIERÍA DE PETRÓLEOS
NEIVA HUILA
2019**

**DETERMINACIÓN DE LA PRESIÓN ONSET EN YACIMIENTOS
PETROLÍFEROS CON REDES NEURONALES IMPLEMENTADAS EN
PYTHON BASADO EN MODELO ESTADÍSTICO.**

**JUAN DANIEL VARGAS PERDOMO
GERMÁN CAMILO ARCE CALDERÓN**

**Trabajo de grado presentado como requisito para optar al Título de
Ingeniero de Petróleos.**

**Director
ING. JAIRO ANTONIO SEPÚLVEDA GAONA, M. Sc.
Codirector
ING. VLADIMIR MOSQUERA CERQUERA, M. Sc.**

**UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
GRUPO DE INVESTIGACIÓN COMPORTAMIENTO DE FASES
PROGRAMA DE INGENIERÍA DE PETRÓLEOS
NEIVA HUILA
2019**

Nota de aceptación

Firma del director

Firma del codirector

Firma del jurado

Firma del jurado

Neiva, Huila Noviembre ___ de 2019

Ante todo, gracias a Dios por guiarme en este arduo camino y permitirme culminar con esta etapa de mi vida, por enseñarme, que para alcanzar mis metas solo se debe tener disciplina y ser perseverante.

A mis padres por apoyarme en todos los momentos y situaciones por los cuales yo atravesé a lo largo de mi carrera, por enseñarme todos los valores que me caracterizan como persona y por ayudarme a tomar decisiones que fueron muy importantes en mi vida.

A mi Hermano, el cual fue una persona crucial en esta etapa universitaria, debido a que me guio y me aconsejó en momentos difíciles. Es mi espejo y modelo a seguir.

A una persona muy especial en mi vida y que influyo positivamente en mi etapa universitaria, que me aconsejo y me ayudo a crecer como persona y como profesional, María Fernanda.

Germán

Primero que todo, dar gracias a Dios por orientarme hacia el camino correcto para poder tomar las mejores decisiones y acompañarme a superar los obstáculos que se van presentando a lo largo a lo largo de la carrera.

A mi familia, por estar siempre apoyándome a lo largo de esta etapa que está por culminar, por brindarme las herramientas necesarias para cumplir con todas las responsabilidades, enseñarme los valores necesarios para crecer cada vez más como persona íntegra. Fueron ellos quienes constantemente me motivaron a alcanzar mis sueños y metas.

Juan

AGRADECIMIENTOS

Los autores del proyecto expresan unos sinceros agradecimientos a:

Los ingenieros Vladimir Mosquera Cerquera y Jairo Antonio Sepúlveda Gaona por brindarnos todo su apoyo y el tiempo necesario para la realización de este trabajo en conjunto con los programas de la Facultad de Ingeniería, en especial ingeniero Vladimir por guiarnos en todo lo relacionado con el software para la implementación de la red neuronal y al ingeniero Jairo Antonio Sepúlveda Gaona por ayudarnos en el análisis e interpretación de los resultados obtenidos en este trabajo de grado ya que es una de las etapas más importantes para la validación de este trabajo.

A nuestros compañeros que con su apoyo incondicional nos motivaron a seguir y avanzar poco a poco en cada etapa del proyecto, en especial a Karen Yulieth Cerón Manrique, Francisco González, María Paula Salazar García y Angela María Palomino.

RESUMEN

TÍTULO:

DETERMINACIÓN DE LA PRESIÓN ONSET EN YACIMIENTOS PETROLÍFEROS CON REDES NEURONALES IMPLEMENTADAS EN PYTHON BASADO EN MODELO ESTADÍSTICO.

AUTORES:

GERMÁN CAMILO ARCE CALDERÓN
JUAN DANIEL VARGAS PERDOMO

PALABRAS CLAVES:

Presión Onset, Red Neuronal, Conexión Completa, Inteligencia Artificial, Python, Temperatura, Presión de Saturación y Máquinas de Soporte Vectorial.

DESCRIPCIÓN:

En este trabajo se desarrolló e implementó una red neuronal totalmente conectada (full connect) la cuál es llamada AsphPy, y un algoritmo de “Machine Learning” con la finalidad de predecir la presión Onset en los yacimientos de petróleo. Esta red neuronal fue elaborada en el lenguaje de programación Python y está basada en funciones estadísticas para el elaboración y evaluación del rendimiento del modelo de predicción. El algoritmo de Machine Learning implementó un soporte de maquina vectorial (SVM, Support Vector Machine), que tuvo como finalidad comparar sus resultados con el rendimiento de la red AsphPy.

Una finalidad importante de este trabajo es el uso de Python en modelos de predicción de la presión Onset realizados por el grupo de investigación COFA. El primer modelo, presentado por Cristian Stiven Loaiza Cano y Samuel Quintero Herrera en el trabajo “Desarrollo de un Modelo para la Determinación de la Presión de Inicio de Precipitación de Asfáltenos (Onset) en Yacimientos de Petróleo”, es la base para la elaboración de AsphPy. El segundo modelo, presentado por Nathalie Flor Olave y Nicolas Hernando Quesada Bahamon en el trabajo “Determinación de la Presión Onset Usando Análisis Estadístico ANOVA en el Modelo de la Red Neuronal” desarrollaron mejoras al poder implementar la red neuronal basándose en un análisis ANOVA; el cual brinda información de gran ayuda para observar qué componentes o datos de entrada son más relevantes para la determinación de la presión Onset.

ABSTRACT

TITLE:

DETERMINATION OF ONSET PRESSURE IN OIL RESERVOIRS WITH NEURAL NETWORKS IMPLEMENTED IN PYTHON BASED ON A STATISTICAL MODEL.

AUTHORS:

GERMÁN CAMILO ARCE CALDERÓN
JUAN DANIEL VARGAS PERDOMO

KEYWORDS:

Onset pressure; artificial neural network; complete connection; artificial intelligence; Python; temperature; saturation pressure; support vector machine.

DESCRIPTION:

In this work, a fully connected neural network (full connect), called AsphPy, and a machine learning algorithm were developed and implemented in order to predict Onset pressure in oil reservoirs. This neural network was developed in the Python programming language and is based on statistics for the evaluation of the performance of the artificial neural network. The Machine Learning algorithm implemented a support Vector Machine (SVM), which was intended to compare its results with the performance of the AsphPy network.

An important purpose of this work is the use of Python in Onset pressure prediction models by the COFA research group. The first model, presented by Cristian Stiven Loaiza Cano and Samuel Quintero Herrera in the work "development of a model for determining the asphaltene onset pressure in oil reservoirs", is the basis for the creation of AsphPy. In the second model, presented by Nathalie Flor Olave and Nicolas Hernando Quesada Bahamon in the work "determination of Onset pressure using ANOVA statistical analysis in neural network model", improvements were developed in the ability to implement the neural network based on an ANOVA analysis; which provided helpful information to see which components or input data are most relevant to the determination of Onset pressure.

CONTENIDO

<u>1. CAPÍTULO UNO: FUNDAMENTOS BÁSICOS</u>	13
1.1. PETRÓLEOS Y ASFALTENOS	13
1.2. ANÁLISIS SARA	13
1.3. PRECIPITACIÓN DE ASFALTENOS	14
1.4. PYTHON	15
1.5. KERAS.....	16
1.6. SCIKIT-LEARN	16
1.7. REDES NEURONALES.....	16
1.8. MÁQUINAS DE SOPORTE VECTORIAL.	17
<u>2. CAPÍTULO DOS: SIMULACIÓN DE MUESTRAS (CMG)</u>	18
2.1. PARÁMETROS DE SIMULACIÓN.....	18
2.2. PROCESO DE SELECCIÓN DE MUESTRAS.....	18
<u>3. CAPÍTULO TRES: RED NEURONAL ARTIFICIAL Y MÁQUINA DE SOPORTE VECTORIAL</u>	25
3.1. LIBRERÍAS.	25
3.2. DE EXCEL A PYTHON.	25
3.3. RED NEURONAL ARTIFICIAL.	26
3.3.1. DATOS DE ENTRENAMIENTO Y DATOS DE PRUEBA.....	26
3.3.2. NORMALIZACIÓN DEL DATASET.....	27
3.3.3. FUNCIÓN DE ACTIVACIÓN (RECTIFIED LINEAR UNIT).	28
3.3.4. ARQUITECTURA.....	29
3.3.5. FUNCIÓN DE PERDIDAS, OPTIMIZADOR Y MÉTRICAS.....	30
3.3.6. ENTRENAMIENTO DE LA RED NEURONAL.....	32
3.4. MÁQUINAS DE SOPORTE VECTORIAL.	32
3.4.1. DATOS PARA LA FORMULACIÓN DEL MODELO.....	32
3.4.2. KERNEL Y MODELO DE PREDICCIÓN SVM.....	33
<u>4. RESULTADOS Y DISCUSIONES</u>	35
4.1. VALIDACIÓN DEL MODELO	35
4.2. ANÁLISIS DEL COMPORTAMIENTO DE LAS MUESTRAS CUANDO SE INYECTA GAS. 43	
4.3. ANÁLISIS DE LAS TENDENCIAS DE LAS CURVAS AOP	44
4.3.1 GRUPO 1	44
4.3.2 GRUPO 2	47
4.3.3. GRUPO 3	49
4.3.4. GRUPO 4	51
<u>5. CONCLUSIONES Y RECOMENDACIONES</u>	53
5.1. CONCLUSIONES.....	53
5.2. RECOMENDACIONES.....	54

BIBLIOGRAFÍA	56
--------------------	----

ANEXOS	58
--------------	----

COMPARACIÓN ENTRE LA PRESIÓN ONSET REAL DE 100 MUESTRAS ALEATORIAS, LOS PREDICHOS CON ASPHPY Y EL SVM.	58
COMPOSICIÓN QUIMICA DE LAS MUESTRAS SIN NORMALIZAR.	60
ANALISIS SARA SIN NORMALIZAR.	62
PRESION ONSET Y PRESION DE SATURACIÓN A 5 TEMPERATURAS.	64
RED NEURONAL Y SUPPORT VECTOR MACHINE CODIGO.	66

LISTA DE TABLAS

TABLA 2.1 NORMALIZACIÓN DE LA COMPOSICIÓN DESPUÉS DE LA INYECCIÓN.....	19
TABLA 4.1 ERROR PROMEDIO DE LOS DOS MODELOS DE PREDICCIÓN.....	35
TABLA 4.2 COMPOSICIÓN QUÍMICA DE LAS 8 MUESTRAS ESCOGIDAS ALEATORIAMENTE.	39
TABLA 4.3 ANÁLISIS SARA Y PRESIÓN DE SATURACIÓN (Pb) DE LAS 8 MUESTRAS ESCOGIDAS ALEATORIAMENTE.	40
TABLA 4.4 ERROR MÁXIMO Y ERROR PROMEDIO CON RESPECTO AL NÚMERO DE ÉPOCAS PARA ENTRENAMIENTO DE LA RED NEURONAL.....	42
TABLA 4.4. CÁLCULO DEL PORCENTAJE DE ERROR EN LAS MUESTRAS SIMULADAS EN CMG	43

LISTA DE FIGURAS

FIGURA 1.1 PRESIÓN ONSET DE PRECIPITACIÓN DE ASFÁLTENOS (CANO Y HERRERA, 2016)	15
FIGURA 2.1 TENDENCIA DE LA INYECCIÓN DE METANO (PEYMAN BAHRAMI, 2015)	18
FIGURA 2.2 TOMA DE DATOS DE PRESIÓN Y TEMPERATURA POR MEDIO DE DIGITIZE IT 2016.....	20
FIGURA 2.3 GRÁFICA OBTENIDA DE LOS PUNTOS EXPORTADOS DE DIGITIZE IT.....	20
FIGURA 2.4 PROPIEDADES NECESARIAS PARA FRACCIONAR.	21
FIGURA 2.5 ENVOLVENTE PRESIÓN VS TEMPERATURA.....	22
FIGURA 2.6 GRÁFICA % ASFÁLTENOS PRECIPITADOS VS PRESIÓN.	24
FIGURA 3.1 LIBRERÍAS DE PYTHON USADAS PARA LA CREACIÓN DE LOS MODELOS DE RED NEURONAL Y MÁQUINAS DE SOPORTE VECTORIAL.	25
FIGURA 3.2 LECTURA DE ARCHIVOS DE EXCEL Y CREACIÓN DE LAS VARIABLES PRINCIPALES.....	26
FIGURA 3.3 IDENTIFICADORES DE ENTRENAMIENTO Y DE PRUEBA.....	26
FIGURA 3.4 MUESTRAS DE ENTRENAMIENTO Y DE PRUEBA.	27
FIGURA 3.5 LIBRERÍA PARA LA NORMALIZACIÓN DEL DATASET.....	27
FIGURA 3.6 DECLARACIÓN DE LA VARIABLE SCALER.	28
FIGURA 3.7 NORMALIZACIÓN DEL CONJUNTO DE DATOS DE ENTRADA.	28
FIGURA 3.8 REPRESENTACIÓN GRÁFICA DE LA FUNCIÓN DE ACTIVACIÓN.	28
FIGURA 3.9 ARQUITECTURA DE LA RED NEURONAL ASPHPY.	29
FIGURA 3.10 DECLARACIÓN DE LAS CAPAS EN LA RED NEURONAL.	30
FIGURA 3.11 FUNCIÓN DE PERDIDAS, OPTIMIZADOR Y MÉTRICA DE EVALUACIÓN PARA LA CREACIÓN DEL MODELO. ...	31
FIGURA 3.12 REPRESENTACIÓN GRÁFICA DE LAS PERDIDAS EN EL MODELO DE PREDICCIÓN.....	32
FIGURA 3.13 LECTURA DE LA BASE DE DATOS EN EXCEL Y DECLARACIÓN DE LAS VARIABLES DE ENTRENAMIENTO. ...	33
FIGURA 3.14 DECLARACIÓN DE PARÁMETROS PARA LAS MÁQUINAS DE SOPORTE VECTORIAL.	34
FIGURA 3.15 DECLARACIÓN DE PARÁMETROS PARA LAS MÁQUINAS DE SOPORTE VECTORIAL.	34
FIGURA 4.1 PRESIÓN VS TEMPERATURA MUESTRA (1) PRESIÓN ONSET (UPPER).	36
FIGURA 4.2 PRESIÓN VS TEMPERATURA MUESTRA (2) PRESIÓN ONSET (UPPER).	36
FIGURA 4.3 PRESIÓN VS TEMPERATURA MUESTRA (3) PRESIÓN ONSET (UPPER).	36
FIGURA 4.4 PRESIÓN VS TEMPERATURA MUESTRA (4) PRESIÓN ONSET (UPPER).	37
FIGURA 4.5 PRESIÓN VS TEMPERATURA MUESTRA (5) PRESIÓN ONSET (UPPER).	37
FIGURA 4.6 PRESIÓN VS TEMPERATURA MUESTRA (6) PRESIÓN ONSET (UPPER).	37
FIGURA 4.7 PRESIÓN VS TEMPERATURA MUESTRA (7) PRESIÓN ONSET (UPPER).	38
FIGURA 4.8 PRESIÓN VS TEMPERATURA MUESTRA (8) PRESIÓN ONSET (UPPER).	38
FIGURA 4.9 PRESIÓN REAL VS PREDICCIONES DE LA RED NEURONAL.	41
FIGURA 4.10 PRESIÓN REAL VS PREDICCIONES DE LAS MÁQUINAS DE SOPORTE VECTORIAL.	41
FIGURA 4.11 PRESIÓN REAL VS PREDICCIONES DE LOS MODELOS DE PREDICCIÓN.	42
FIGURA 4.12 GRÁFICA AOP VS TEMPERATURA GRUPO 1.....	45
FIGURA 4.13 TENDENCIA DE LA PRESIÓN DE SATURACIÓN DEL GRUPO 1 CON RESPECTO A LA VARIACIÓN DE TEMPERATURA.	45
FIGURA 4.14 ANÁLISIS SARA DEL GRUPO 1.....	46
FIGURA 4.15 ANÁLISIS CROMATOGRÁFICO DEL GRUPO 1.....	46
FIGURA 4.16 GRÁFICA AOP VS TEMPERATURA GRUPO 2.....	47
FIGURA 4.17 TENDENCIA DE LA PRESIÓN DE SATURACIÓN DEL GRUPO 2 CON RESPECTO A LA VARIACIÓN DE TEMPERATURA.	47
FIGURA 4.18 ANÁLISIS SARA DEL GRUPO 2.....	48
FIGURA 4.19 ANÁLISIS CROMATOGRÁFICO DEL GRUPO 2.....	48
FIGURA 4.20 GRÁFICA AOP VS TEMPERATURA GRUPO 3.	49
FIGURA 4.21 TENDENCIA DE LA PRESIÓN DE SATURACIÓN DEL GRUPO 3 CON RESPECTO A LA VARIACIÓN DE TEMPERATURA.	49
FIGURA 4.22 ANÁLISIS CROMATOGRÁFICO DEL GRUPO 3.....	50
FIGURA 4.23 ANÁLISIS SARA DEL GRUPO 3.....	50
FIGURA 4.24 GRÁFICA AOP VS TEMPERATURA GRUPO 4.....	51
FIGURA 4.25 TENDENCIA DE LA PRESIÓN DE SATURACIÓN DEL GRUPO 4 CON RESPECTO A LA VARIACIÓN DE TEMPERATURA.	51
FIGURA 4.26 ANÁLISIS CROMATOGRÁFICO DEL GRUPO 4.....	52
FIGURA 4.27 ANÁLISIS SARA DEL GRUPO 4.....	52

1. CAPÍTULO UNO: FUNDAMENTOS BÁSICOS

1.1. PETRÓLEOS Y ASFALTENOS

El petróleo es una mezcla de compuestos químicos orgánicos llamados hidrocarburos, término que hace referencia a los compuestos que están formados por carbono e hidrógeno. Se puede clasificar de acuerdo con su densidad por medio de una escala arbitraria dividida en grados API según el Instituto Americano del Petróleo. Un petróleo se considera liviano si está en un rango de 30-40°, petróleo mediano en el intervalo 22-29.9°, pesado entre 10-21.9° y extrapesado si es menor a 10°. También está compuesto por compuestos no orgánicos como el azufre, nitrógeno, oxígeno y pequeñas trazas de metales como el hierro, el zinc y níquel (Pérez, 2015). Los compuestos no orgánicos son considerados como impurezas, estas pueden aparecer en cualquier intervalo de ebullición del aceite, pero por lo general se concentran en las fracciones más pesadas, y están separados en: compuestos de azufre, compuestos de nitrógeno, compuestos oxigenados, compuestos de metales, resinas y, por último, los asfaltenos, tema de interés en este trabajo (Padilla D. y Watt K., 2012).

La depositación de asfaltenos se manifiesta en casi todos los aspectos de producción, procesamiento y transporte del petróleo. La precipitación de asfaltenos en los yacimientos, pozos, y facilidades tiene un efecto negativo en la economía debido a la reducción de la productividad del pozo. La naturaleza y comportamiento de los asfaltenos en crudos es complicada, estos son pensados como estructuras dispuestas en anillos múltiples apilados. Consisten principalmente de carbono, hidrógeno y una pequeña proporción de heteroelementos como oxígeno, azufre y nitrógeno (E. F. Gholoum, G. P. Oskui y M. Salman, 2003).

Los asfaltenos son definidos como la fracción polar, poliaromática y más pesada del petróleo, que generalmente son caracterizados por ser insolubles en n-heptano o n-pentano (E. F. Gholoum, G. P. Oskui y M. Salman, 2003). Pero si son solubles en aromáticos como el benceno y tolueno. Los asfaltenos no tienen un punto de fusión definido y usualmente espuman y expanden cuando se calientan y pueden dejar residuos carbonáceos, el peso molecular de los asfaltenos varía entre 500 hasta 10000 g/mol y el punto de ebullición está por encima de 540°C (Paredes, 2013).

1.2. ANÁLISIS SARA

El contenido de asfaltenos es un factor muy importante a la hora de determinar las propiedades del petróleo. Existe un método de laboratorio llamado SARA, mediante el cual se ha podido cuantificar la fracción de asfaltenos. Esta técnica separa el petróleo muerto, es decir, el crudo sin componentes gaseosos, en saturados, aromáticos, resinas y asfaltenos, dependiendo de la solubilidad y

polaridad de cada uno de ellos. (Al-Qasim, A., y Bubshait, A, 2017). La Figura 1.1 muestra una representación esquemática de este método.

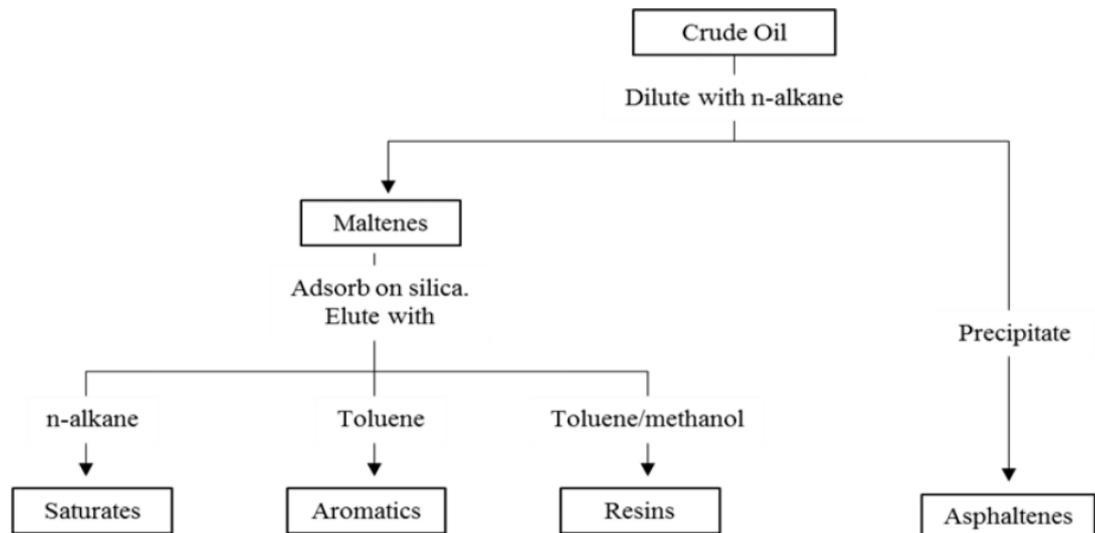


Figura 1.1. Esquema típico en separación de hidrocarburos en fracciones SARA (Al-Qasim, A., y Bubshait, A, 2017).

En el análisis SARA la cromatografía líquida está precedida por la precipitación de los asfaltenos usando como disolvente un n-alcano. Los componentes remanentes, los máltenos, se separan por cromatografía de elución, haciendo pasar la muestra a través de una columna. Cada componente se remueve de la columna con la ayuda de diferentes fases estacionarias, lavándolos con diversos solventes de polaridad variada. (Fernández Salgado y Yatzirih, 2017).

1.3. PRECIPITACIÓN DE ASFALTENOS

Precipitación es el proceso mediante el cual los compuestos de este origen aparecen en fase insoluble, como resultado de alteraciones ocurridas en las condiciones del medio, ya sean termodinámicas u operacionales. Dando como resultado la desestabilización del material asfaltenico, que se manifiesta cuando las partículas caen, rompen la emulsión, chocan y se agregan unas a otras. Una vez estando en esa etapa, los flóculos se mantienen suspendidos sin generar problemas. Cuando hay un aumento de los flóculos, estos forman agregados que pueden permanecer suspendidos para posteriormente ser arrastrados por la corriente del fluido si la velocidad es alta (Padilla D. y Watt K., 2012). Por último, ocurre la deposición, proceso de intercambio de flóculos entre la fase oleica y la superficie de la roca o la tubería, dependiendo del medio en donde se encuentre (Al-Qasim, A., y Bubshait, A, 2017).

Una de las técnicas más populares para evaluar la precipitación de asfaltenos consiste en la determinación de la presión de precipitación de asfaltenos (AOP), parámetro importante a la hora de modelar con asfaltenos. Esta variable es

definida como la presión a la cual la fracción del petróleo alcanza el límite que conlleva a la floculación y precipitación de asfáltenos. (Padilla D. y Watt K., 2012). Todo el asfalteno precipitado no se deposita ni causa daño a la formación; los precipitados pueden ser arrastrados por el fluido, pueden depositarse en la superficie de la roca, o se agregan hasta formar un precipitado mayor que logre taponar las gargantas de poro. La figura 1.2 muestra la gráfica que representa la presión Onset de precipitación de asfáltenos, en esta se observa que a medida que la presión disminuye, la cantidad de precipitado aumenta, hasta un máximo en la presión de saturación (presión de burbuja) y por debajo la cantidad de precipitado disminuye debido a que el gas que está en solución es liberado causando que los asfáltenos se estabilicen y por lo tanto se disuelvan nuevamente en el fluido de yacimiento.

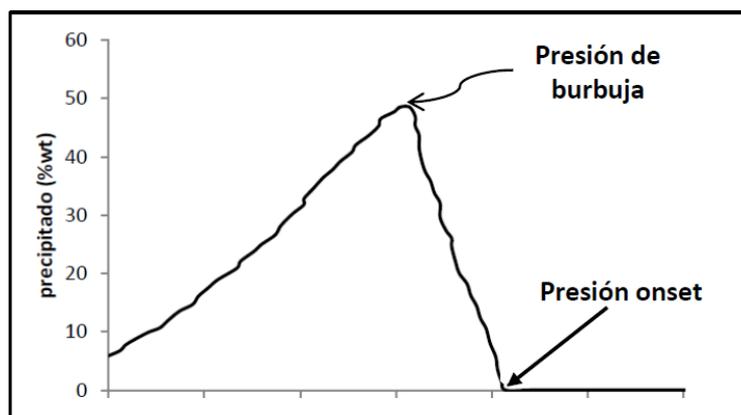


Figura 1.1 Presión Onset de precipitación de asfáltenos (Cano y Herrera, 2016)

El efecto de la presión es más pronunciado en cercanías al punto de burbuja. Los asfáltenos se estabilizan a altas presiones hasta el punto de presión *Onset*, en donde la concentración de asfáltenos se hace cero. Al disminuir la presión durante el proceso de producción de petróleo, el volumen relativo de la fracción de componentes ligeros aumenta y la solubilidad del asfalteno en el crudo disminuye, y empieza el proceso de floculación (Pérez, 2015).

1.4. PYTHON

Python es un lenguaje de programación libre (Open Source) catalogado como un lenguaje de alto nivel. Sumado a esto tiene grandes ventajas como Una sintaxis simple, multiplataforma disponible para Linux, Windows y Mac, entre otras.

Especialmente Python, cuenta con una gran cantidad de librerías para el desarrollo de trabajos en modelos de predicción o clasificación, manipulación de datos, operaciones matemáticas, visualización o representación gráfica de los datos, entre otros tipos de procesos. Debido a esto ha provocado que en los últimos años su implementación y desarrollo haya crecido radicalmente comparado con otros lenguajes de programación.

Este lenguaje fue creado por Guido Van Rossum a inicios de los años 90's con la finalidad de desarrollar un lenguaje para la comunidad que quería incursionar en el mundo de la programación, desarrollando un lenguaje compacto y estructurado, a su vez hacía que su sintaxis fuera legible. En contraste, Python fue desarrollado como una segunda oportunidad que se le dio al lenguaje de programación ABC, el cual no evoluciono debido a las limitaciones tecnológicas de la época.

1.5. KERAS

Keras es un API diseñada e implementada en Python; que tiene como objetivo facilitar la experiencia al usuario al momento de entrar en el mundo de la inteligencia artificial. Esta librería puede ofrecer una gran cantidad de herramientas para la construcción de una red neuronal, es decir; el usuario tiene la oportunidad de poder variar los parámetros de la red que desea implementar, desde la arquitectura de la red, la función de costo, la función de activación, entre otros parámetros.

1.6. SCIKIT-LEARN

Scikit-Learn es una librería para el desarrollo de modelos de predicción o clasificación basándose en algoritmos de Machine Learning. Los algoritmos más relevantes en esta librería son: Máquina de soporte vectorial, K-means, bosques aleatorios, regresión lineal y regresión logística. Es una de las librerías o APIs más implementadas en Python, debido a su gran versatilidad y variedad de modelos, además de los algoritmos que se puede implementar, Scikit Learn tiene herramientas útiles en el mundo del análisis de datos como reducción de dimensiones en los dataset, preparación de datos de prueba y entrenamiento, optimizador de parámetros, limpieza de datos, etc.

1.7. REDES NEURONALES.

El hombre se ha caracterizado siempre por su búsqueda constante de nuevas vías para mejorar sus condiciones de vida. Estos esfuerzos le han servido para reducir el trabajo en aquellas operaciones en las que la fuerza juega un papel primordial. Los progresos obtenidos han permitido dirigir estos esfuerzos a otros campos, como, por ejemplo, a la construcción de máquinas calculadoras que ayuden a resolver de forma automática y rápida determinadas operaciones que resultan tediosas cuando se realizan a mano (Ruiz, Basualdo, y Matich, 2001).

Por esto es por lo que el hombre al hallar las dificultades que implican la realización de algunas operaciones viéndose afectado, asume una tarea bastante compleja, la cual le permitiera encontrar una herramienta para reducir de manera tangencial las problemáticas que se puedan encontrar en las operaciones, emprendiendo el análisis de la mente humana.

Las redes neuronales son más que otra forma de emular ciertas características propias de los humanos, como la capacidad de memorizar y de asociar hechos. Si se examinan con atención aquellos problemas que no pueden expresarse a través de un algoritmo, se observará que todos ellos tienen una característica en común: la experiencia. El hombre es capaz de resolver estas situaciones acudiendo a la experiencia acumulada. Así, parece claro que una forma de aproximarse al problema consista en la construcción de sistemas que sean capaces de reproducir esta característica humana. En definitiva, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el ejemplo más perfecto del que disponemos para un sistema que es capaz de adquirir conocimiento a través de la experiencia. Una red neuronal es “un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: *la neurona*” (Matich, 2001).

1.8. MÁQUINAS DE SOPORTE VECTORIAL.

Este algoritmo de Machine Learning es muy implementado en la industria para casos de estudio tanto de regresión como de clasificación, es considerado como un algoritmo supervisado visto a que se necesitan datos de entrada y sus respectivos datos de salida. Las máquinas de soporte vectorial pueden también adaptarse para resolver problemas de regresión. En estos casos, es muy común designarlas por el acrónimo SVR (del inglés Support Vector Regression) (Súarez, 2013).

En sus inicios, este algoritmo de Machine Learning se desarrolló solo para casos de estudios en los cuales se hiciera una clasificación binaria, es decir, es o no es. A medida que avanzaba y evolucionaba este algoritmo de Machine Learning su campo de aplicación se fue ampliando, empezando con clasificación multiclase, que consistía en identificar qué cualidades o patrones tenían los datos y así mismo asociarlos con un grupo de interés. Otra aplicación fue la regresión, en otras palabras, este algoritmo puede predecir o calcular cantidades partiendo de unos datos de entrada; para esto, este algoritmo cuenta con kernel o una función que nos ayuda a predecir y calcular una función en términos de los datos de entrada, concretamente encontrar una relación entre estos datos.

2. CAPÍTULO DOS: SIMULACIÓN DE MUESTRAS (CMG)

2.1. PARÁMETROS DE SIMULACIÓN.

El proceso más importante en este proyecto es el completamiento de los datos, en consecuencia, se debe tener todos los parámetros que se consideran importantes o que influyen en la precipitación de los asfaltenos. Es por esto por lo que se consideraron como parámetros importantes al momento de predecir la presión Onset las siguientes variables:

- Composición química de la muestra de petróleo.
- Análisis SARA (Saturados, Aromáticos, Resinas, Asfaltenos).
- Presión de Burbuja (P_b)

2.2. PROCESO DE SELECCIÓN DE MUESTRAS

Las curvas de las muestras que se tenían anteriormente, es decir, de las versiones anteriores de este estudio, tenían un comportamiento descendente en donde la presión onset disminuía con el aumento de la temperatura. A raíz de eso, se decide buscar otros comportamientos de estas curvas en diferentes estudios, principalmente aquellos en donde se inyectaba gas con el objetivo de aumentar el recobro de petróleo. Para estos casos, en dichos trabajos de investigación, se evalúa la influencia al suministrar CO_2 , C_1 , N_2 y gas del mismo yacimiento (gas asociado) a diferentes concentraciones. En la figura 2.1 se ilustra el comportamiento que tiene la presión de burbuja y la presión onset de la muestra, al inyectarle metano en diferentes proporciones (5% y 10%), incluyendo una composición inicial del yacimiento sin alteraciones.

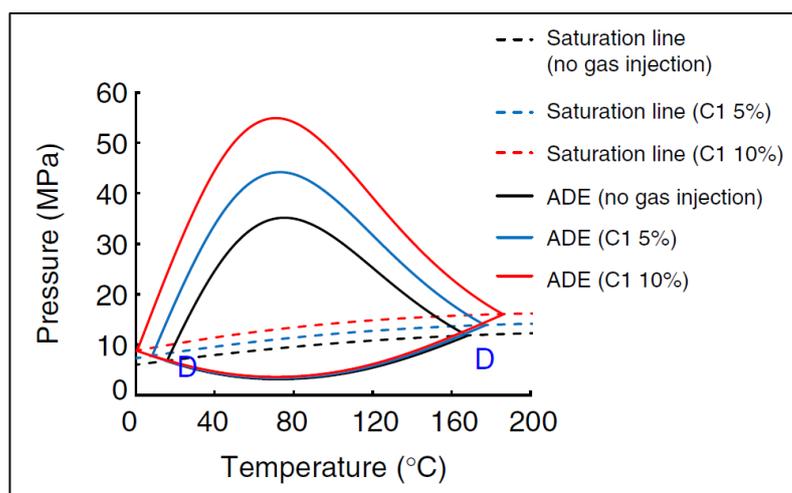


Figura 2.1 Tendencia de la inyección de metano (Peyman Bahrami, 2015)

De todos estos estudios se obtenían datos experimentales en donde se construían curvas evidenciando la presión de burbuja (Pb) y la presión onset de los asfaltenos (AOP) a diferentes temperaturas, teniendo en cuenta la concentración de los fluidos que son inyectados en el yacimiento para mejorar el recobro de petróleo.

Una vez identificadas las curvas, se seleccionaban las que cumplieran con el rango de temperatura que se requiere para este trabajo (320 - 400) k. Luego la muestra original de yacimiento se tuvo que normalizar al inyectarle determinada concentración de fluido para poder actualizar las fracciones molares de la mezcla. Ver tabla 2.1.

Tabla 2.1 Normalización de la composición después de la inyección.

COMPONENTES	MW	Densidad	Muestra de yacimiento	Inyección C1	
	(g/mol)	(g/cm3)	Composición Inicial	Para 5%	Para 10%
				Composición mezcla	
NITROGENO	28.04	0.8090	0.39	0.371	0.351
DIOXIDO DE CARBONO	44.01	0.8170	1.74	1.653	1.566
METANO	16.04	0.3000	20.55	24.523	28.495
ETANO	30.07	0.3560	7.31	6.945	6.579
PROPANO	44.10	0.5080	5.34	5.073	4.806
ISOBUTANO	58.12	0.5670	1.00	0.950	0.900
N-BUTANO	58.12	0.5860	3.65	3.468	3.285
ISOPENTANO	72.15	0.6250	3.10	2.945	2.790
N-PENTANO	72.15	0.6310	4.75	4.513	4.275
N-HEXANO	84.00	0.6900	5.48	5.206	4.932
C7	96.00	0.7270	3.23	3.069	2.907
C8	107.00	0.7490	1.32	1.254	1.188
C9	121.00	0.7680	2.27	2.157	2.043
C10	134.00	0.7820	2.19	2.081	1.971
C11	147.00	0.7930	1.81	1.720	1.629
C12+	370.00	0.9769	35.87	34.077	32.283
Total	-	-	100.00	100.000	100.000

Después mediante el programa “Digitize it 2016”, se procedió a seleccionar puntos en dichas curvas, para obtener coordenadas x-y (temperatura vs Pb / AOP) según corresponda, tomando como referencia dos puntos para ajustar la escala observar la Figura 2.2. Los datos obtenidos fueron exportados a Excel en donde se graficaban y se obtenía nuevamente la curva, para compararla con la de cada caso de estudio, de allí se obtenían los valores de presión. De la gráfica resultante, se le agrega una línea de tendencia con el mejor tipo de polinomio que ajuste, en donde se reemplaza la “x” para las cinco temperaturas (46.85,

66.85, 86.85, 106.85, 126.85) °C. Para este caso, el polinomio que mejor ajuste fue de cuarto grado ver Figura 2.3.

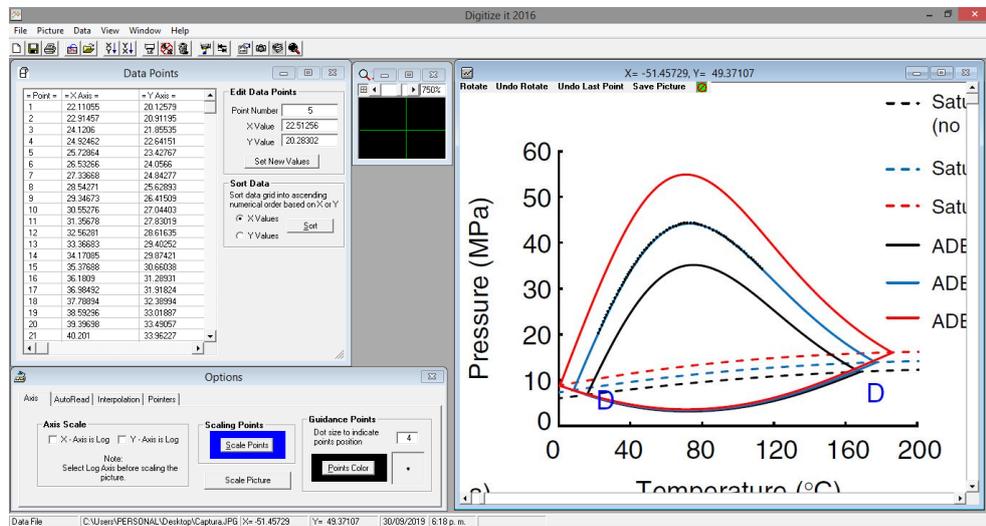


Figura 2.2 Toma de datos de presión y temperatura por medio de Digitize It 2016

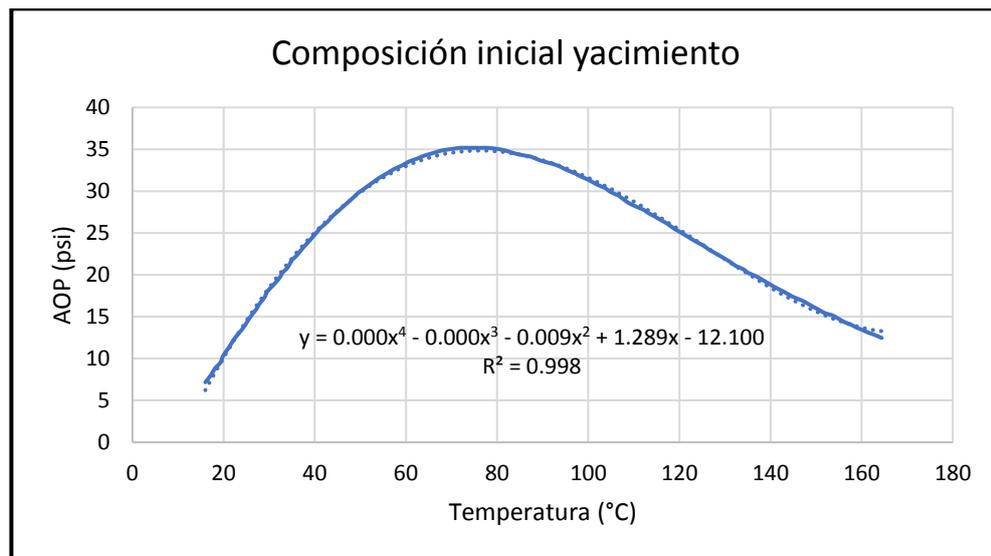


Figura 2.3 Gráfica obtenida de los puntos exportados de Digitize It.

Con todos los datos establecidos, se debe realizar una comprobación de estos datos por medio de la plataforma “CMG”, programa WinProp. Una vez estando allí, se ingresan los componentes con su respectiva fracción molar. Para el caso de la fracción plus, se requiere de ciertas propiedades que deben estar incluidas en el artículo desarrollado por Peyman Bahrami, obsérvese la Figura 2.4.

General	Distribution	Sample 1	Mud Info.
Plus Fraction Property			
MW:	<input type="text" value="370"/>	Sample Name:	<input type="text" value="Sample 1"/>
SG:	<input type="text" value="0.9769"/>	Slope:	<input type="text"/>
Mole Frac:	<input type="text" value="0.3587"/>	No. of SCN Fractions:	<input type="text" value="0"/>
<input type="checkbox"/> Conduct contaminated sample analysis Level of Contamination (weight fraction of mud): <input type="text"/>			

Figura 2.4 Propiedades necesarias para fraccionar.

Además de eso, WinProp cuenta con unos parámetros internos para poder ajustar el número de pseudocomponentes, el rango de peso molecular en el que se establezca y el último componente que para este caso sería el C31+, diversos estudios muestran que el peso de ese componente corresponde al promedio de los asfaltenos.

Cuando no se cuentan con esas propiedades de la fracción plus, se procede a usar una serie de ecuaciones como las que se tienen a continuación:

Una de las mayores dificultades con respecto al ingreso de la composición de los diferentes crudos, fue conocer las propiedades de la fracción plus (Mw, SG y fracción molar). Cuando no se conoce el peso molecular de esa fracción, se procede a calcular mediante la Ecuación 2.1 teniendo en cuenta los componentes no hidrocarburos (CO₂, H₂S y N₂).

$$MW_{C7+} = \frac{MW_{C7+} - \sum_{i=C1}^{C6} MW_i * Z_i}{Z_{C7+}} \quad \text{Ecuación 2.1.}$$

Con respecto a la gravedad específica, se procede a convertir la fracción molar en fracción volumétrica por medio de la Ecuación 2.2:

$$X_i = Z_i * \frac{SG_{crudo}}{SG_i} * \frac{MW_{crudo}}{MW_i} \quad \text{Ecuación 2.2.}$$

$$\sum_{i=1}^7 X_i = 1 \quad \text{Ecuación 2.3.}$$

$$SG_{C7+} = \frac{SG_{C7+} - \sum_{i=C1}^{C6} SG_i * X_i}{X_{C7+}} \quad \text{Ecuación 2.4}$$

siendo: Z_{C7+} la fracción molar
 X_{C7+} la fracción volumétrica.

Después se crea una regresión para ajustar la muestra a las condiciones del estudio y se incluye dentro de ese bloque a la presión de saturación experimental hallada en la investigación. A lo anterior se le conoce como ajuste de las fases liquido-vapor. En la Figura 2.5 se puede observar la envolvente en un gráfico Presión Vs Temperatura, en donde coincide con el dato experimental.

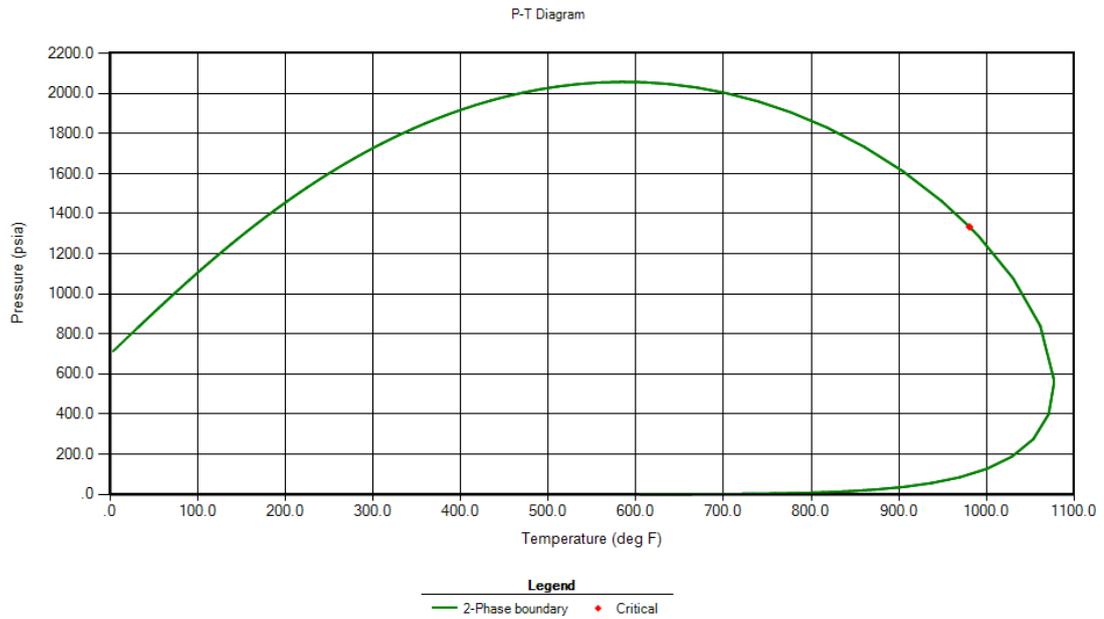


Figura 2.5 Envolvente Presión Vs Temperatura

Luego se agregan las especificaciones de un nuevo componente, para ello se divide la fracción molar del ultimo pseudocomponente en dos, la primera de ellas corresponde a las resinas (C_{31A+}) y la última a los asfaltenos (C_{31B+}). Para calcular la fracción más pesada se usa la Ecuación 2.5.

$$X_{C_{31B+}} = \frac{W_{C_{31B+}} * MW_{oil}}{MW_{C_{31B+}}} \quad \text{Ec. 2.5.}$$

En donde: $W_{C_{31B+}}$ es el porcentaje en peso de asfaltenos, que se obtiene del análisis SARA.

MW_{oil} es el peso molecular de toda la muestra, algunas veces el artículo suministra la información, de no ser así, se corre la simulación en WinProp para obtener el dato.

$MW_{C_{31B+}}$ es el peso molecular de los asfaltenos, obtenido en la corrida de la simulación una vez se hayan cargado las composiciones.

Al bloque de regresión elaborado anteriormente, se le agrega otro nivel correspondiente a la prueba de separador, en donde se incluyen los datos de P

y T de saturación, °API y si hay disponible algún otro dato experimental como el factor volumétrico, viscosidad, GOR, etc.

Seguidamente se ajusta la fase solida-liquida, por tanto, se agregan dos nuevos módulos de composición en donde se especifica la referencia de fugacidad para el modelo de asfaltenos. Para ello se calcula corriendo el programa, lo que establecerá la fugacidad de referencia igual a la fugacidad del componente precipitante en la fase líquida calculada por la ecuación de estado. Importante tener en cuenta que el volumen molar sólido que se calculó anteriormente debe establecerse en un valor ligeramente superior al volumen molar para el componente precipitante predicho por la ecuación de estado. En este punto, se ejecuta el conjunto de datos para verificar la regresión en los datos de PVT del fluido y ver el volumen molar sólido predicho por la EOS (Ltd., 2015)

Posteriormente se procede a graficar, pero antes de obtener se deben ajustar los parámetros del modelo sólido, como el modelo de precipitación sólida utilizado en WinProp es termodinámico, en lugar de cinético, es posible la reversibilidad de la precipitación; es decir, los sólidos precipitados pueden redisolverse en la fase líquida. Este fenómeno se ha observado en el laboratorio para el agotamiento de la presión. Por lo general, la cantidad máxima de precipitación ocurre alrededor de la presión de burbuja del fluido. Por debajo de esta presión, la liberación de gas del aceite cambia el parámetro de solubilidad de la fase líquida y permite la redisolución de los asfaltenos precipitados. Es posible que todo el asfaltenos precipitados vuelva a la solución a presiones suficientemente bajas (Ltd., 2015).

Los parámetros que controlan este comportamiento en el modelo sólido son el volumen molar sólido y el parámetro de interacción entre el componente precipitante y los extremos ligeros del aceite. Aumentar el volumen molar sólido aumenta la cantidad máxima de precipitación a la presión de saturación. El aumento del parámetro de interacción con los extremos de la luz obligará al asfaltenos a redisolverse a presiones más bajas (Ltd., 2015).

Finalmente se genera la gráfica %wt asfaltenos vs P y ahí se puede observar dos puntos clave: Por un lado, la presión de precipitación de asfaltenos, cuya característica se identifica porque a partir de allí comienzan a precipitarse los asfaltenos, tomando como referencia una declinación normal de la presión en un yacimiento cualquiera como es de esperarse. Por otro lado, el pico máximo de la curva representa la presión de burbuja.

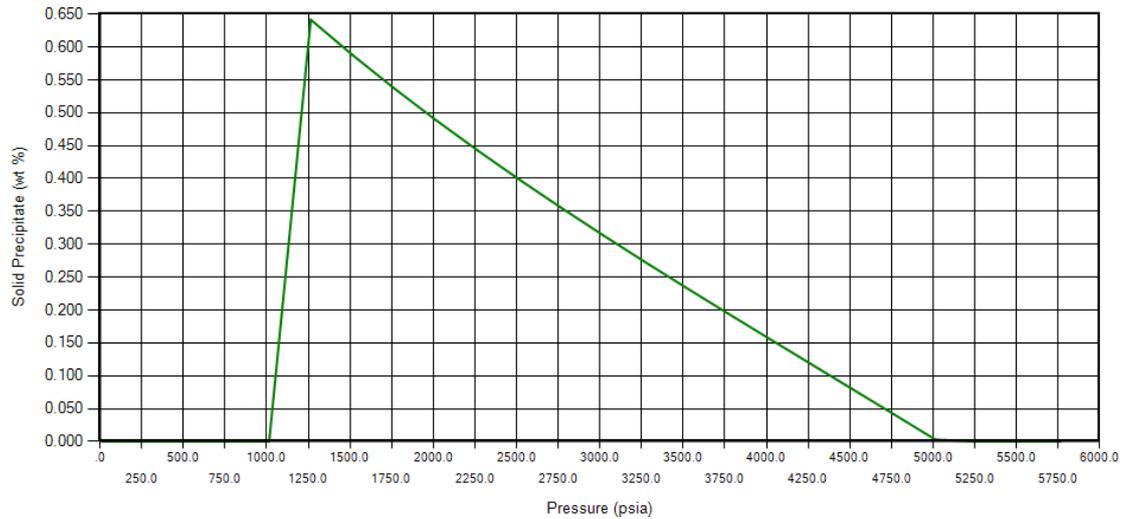


Figura 2.6 Gráfica % Asfaltenos precipitados Vs Presión.

La anterior gráfica, muestra los siguientes resultados para una temperatura de 152.33 °F (340 K):

- La presión de burbuja es igual a 1267 psi
- La presión onset de asfaltenos es igual a 5016 psi.

Con los datos anteriores obtenidos, se calcula el error mediante la ecuación 2.5 comparando los datos experimentales de cada estudio y la simulación en CMG.

$$\%Error = \frac{|P_{CMG} - P_{experimental}|}{P_{experimental}} * 100 \quad \text{Ecuación 2.5}$$

Siendo P la presión onset o la presión de burbuja para cada caso.

Nota: Importante tener en cuenta que el valor de presiones de cada estudio se tomó como dato de referencia (dato real) a la hora de calcular el porcentaje de error, pues las simulaciones en CMG dependen de muchos parámetros y en el bloque de regresión los ajustes la mayoría de los casos se hace por ensayo y error observando cual aporta mejor curva de %Wt asph Vs Presión (forma más parecida a un triángulo).

3. CAPÍTULO TRES: RED NEURONAL ARTIFICIAL Y MÁQUINA DE SOPORTE VECTORIAL.

3.1. LIBRERÍAS.

Para la creación de los dos modelos de la red neuronal y las máquinas de soporte vectorial fue necesario realizar las siguientes operaciones: importar el archivo que contiene las muestras de curdos de Excel a Python, la visualización de los datos y finalmente la formulación de los modelos. En la Figura 3.1 se observa que librerías se utilizaron y como se importaron a Python.

```
from keras.models import Sequential
from keras.layers.core import Dense
from keras import metrics
from sklearn.svm import SVR
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
import pandas as pd
from pandas import ExcelWriter
import numpy as np
```

Figura 3.1 Librerías de Python usadas para la creación de los modelos de red neuronal y máquinas de soporte vectorial.

3.2. DE EXCEL A PYTHON.

La base de datos que almacena todas las muestras de crudo con sus respectivas variables, es decir, la composición química, la temperatura, el análisis SARA, la presión de saturación y la presión onset se realizó en Excel. Mediante la librería “*pandas*” se realizó la exportación de dichas muestras y así se logró hacer la manipulación necesaria para la creación de los modelos de predicción. La lectura del archivo que contiene las muestras de crudo fue posible gracias a la función “*read_excel*”, dicha función tiene como argumento de entrada la dirección en la cual se encuentra ubicado el archivo a leer. Una vez exportado los datos a Python se crearon tres variables, la variable “*predictoras*” almacenó la composición química, la temperatura, el análisis SARA y la presión de saturación, la variable “*ponset*” solo almaceno los valores de la presión onset de cada muestra y finalmente la variable “*Muestras*” estuvo definida mediante la concatenación de las dos variables declaradas anteriormente como se observa en la Figura 3.2.

```
predictoras = pd.read_excel("../datasets/predictoras.xlsx")
ponset = pd.read_excel("../datasets/ponset.xlsx")
Muestras = pd.concat([predictoras,ponset], axis = 1)
```

Figura 3.2 Lectura de archivos de Excel y creación de las variables principales.

3.3. RED NEURONAL ARTIFICIAL.

3.3.1. Datos de entrenamiento y datos de prueba.

Al momento de realizar la construcción del modelo de AsphPy, se debe manipular la base de datos que contiene todas las muestras de crudos con sus respectivas variables predictoras y su variable de interés.

Debido a esto, la base de datos que contiene dichas muestras de crudo se debe dividir en dos subconjuntos cada uno con una finalidad en específico. El primer subconjunto de muestras es definido aleatoriamente gracias a la librería Numpy; dicho subconjunto albergo el 80% de las muestras de la base de datos, estos datos son llamados subconjunto de entrenamiento, por consiguiente, estos datos ayudaran a realizar el modelo de la red AsphPy. El segundo subconjunto de datos contiene el restante de muestras de la base de datos, dichas muestras ayudaran a validar el modelo realizado con el subconjunto de entrenamiento, por lo tanto, reciben el nombre de subconjunto prueba o de test.

Identificadores de entrenamiento y de prueba.

Antes de dividir la base de datos en dos subconjuntos, se debe crear los identificadores aleatorios de las muestras de entrenamiento y las de prueba. Usando las líneas de código mostradas en la Figura 3.3.

```
train_id = np.random.choice(len(X_vals),round(len(X_vals)*0.8), replace=False)
test_id= np.array(list(set(range(len(X_vals)))-set(train_id)))
```

Figura 3.3 Identificadores de entrenamiento y de prueba.

Para obtener los identificadores de entrenamiento se creó una variable “*train_id*”, a esta variable se le realizó una operación gracias a una función que contiene la librería de Numpy. La función que se implementó en esta variable es “*random.choice*”, dicha función permitió crear aleatoriamente los identificadores. El tamaño de esta variable es igual al 80% del número de muestra que contenga la base de datos.

En el caso de los identificadores del subconjunto de prueba, se definió una variable llamada “*test_id*”, dicha variable fue calculada restando los identificadores de la base de datos con los identificadores del subconjunto de entrenamientos. El tamaño de “*test_id*” es del 20% del tamaño de la base de datos, con esto la suma entre los dos subconjuntos dará como resultado el tamaño total de la base de datos.

Selección de las muestras para los subconjuntos.

Creado los identificadores de cada subconjunto se procedió a crear las variables que almacenaron las muestras de crudos, esto es ilustrado la Figura 3.4.

```
x_vals_train = X_vals[train_id]
y_vals_train = Y_vals[train_id]
x_vals_test = X_vals[test_id]
y_vals_test = Y_vals[test_id]
```

Figura 3.4 Muestras de entrenamiento y de prueba.

Siendo “*X_vals*” el array que almacenó cada muestra con su composición química, el análisis SARA y su presión de saturación, por otra parte, la variable “*Y_vals*” solo almacena la Ponset de cada muestra.

3.3.2. Normalización del dataset.

Una vez se definieron los dos subconjuntos que se emplearon para la construcción y validación del modelo para la red neuronal, fue necesario normalizar los valores de cada variable predictora en este caso; la composición química de la muestra, el análisis SARA y la presión de burbuja Pb. Esto tiene como finalidad que ninguna variable predictora predomine con respecto a la variable de interés es decir la Ponset.

Para realizar esta operación se implementó una función de la librería de Scikit-Learn; dicha función es “*MinMaxScaler*”, su finalidad fue calcular el máximo y el mínimo en cada columna en el dataset y así poder normalizar los datos. De la Figura 3.5 a la Figura 3.7, se ilustra su implementación en Python.

Primero se importó dicha función de la librería.

```
from sklearn.preprocessing import MinMaxScaler
```

Figura 3.5 Librería para la normalización del dataset.

Posteriormente se procedió a declarar la variable scaler que contendrá la función *MinMaxScaler()*.

```
scaler = MinMaxScaler()
```

Figura 3.6 Declaración de la variable scaler.

Finalmente, creada la variable o el objeto scaler se implementó la función para normalizar los datos de las variables predictoras.

```
X_vals = scaler.fit_transform(X_vals)
```

Figura 3.7 Normalización del conjunto de datos de entrada.

3.3.3. Función de activación (Rectified Linear Unit).

La función de activación es el parámetro más importante a la hora de implementar una red neuronal; dicha función le agrega la no-linealidad a los datos cuando pasan por cada grafo de la red. Para la implementación de AsphPy la función encargada de introducir la no-linealidad es la función ReLU (Rectified Linear Unit) y es ilustrada en la Ecuación 3.1.

$$F_{(x)} = \begin{cases} 0 & \text{para } x < 0 \\ x & \text{para } x \geq 0 \end{cases} \quad \text{Ecuación 3.1}$$

Como se observó anteriormente, la función de activación $F(x)$ es igual a cero cuando x es menor que cero y la función es igual a x cuando x es igual o mayor a cero. Para tener una mejor comprensión sobre esta función de activación se ilustrará esta función gráficamente en la Figura 3.8.

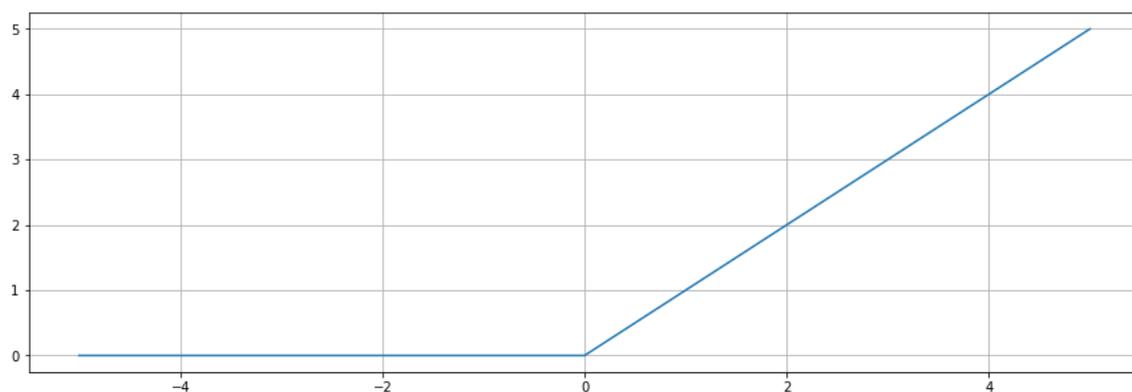


Figura 3.8 Representación gráfica de la función de activación.

3.3.4. Arquitectura.

La arquitectura de la red neuronal artificial es una red totalmente conectada o conocido en inglés como full connected; este tipo de arquitectura consiste, en que cada grafo de una capa está conectado con todos los grafos de la siguiente capa, así es posible conseguir todas las conexiones hacia el grafo de la capa de salida de la red neuronal, en otras palabras, si se hace una analogía con el cerebro humano, este tipo de arquitectura nos expresa que todas las neuronas hacen sinapsis entre ellas.

Sinapsis está dada por la siguiente definición; la sinapsis es un espacio, que hay entre una neurona y otra célula (neurona o no). Un lugar muy activo en el que continuamente suceden cosas. Físicamente es una separación, funcionalmente una conexión que transfiere la información de una célula a otra (Sinapsis, 2014).

La topología de AsphPy es ilustrada en la Figura 3.9, se observa que cuenta con tres etapas. La primera etapa de la red consta de una capa de entrada de datos y contiene N números de grafos, cada grafo tiene como función introducir no-linealidad mediante la función de activación. La segunda etapa de AsphPy cuenta con N número de capas; esta etapa recibe el nombre de capas ocultas o intermedias, cada capa oculta cuenta con N número de grafos en el cual cada grafo realiza su operación mediante la función de activación. Por último, AsphPy cuenta con una etapa de salida, esta etapa contiene solo una capa y un solo grafo. Esta arquitectura se realizó mediante la librería de Python Karas.

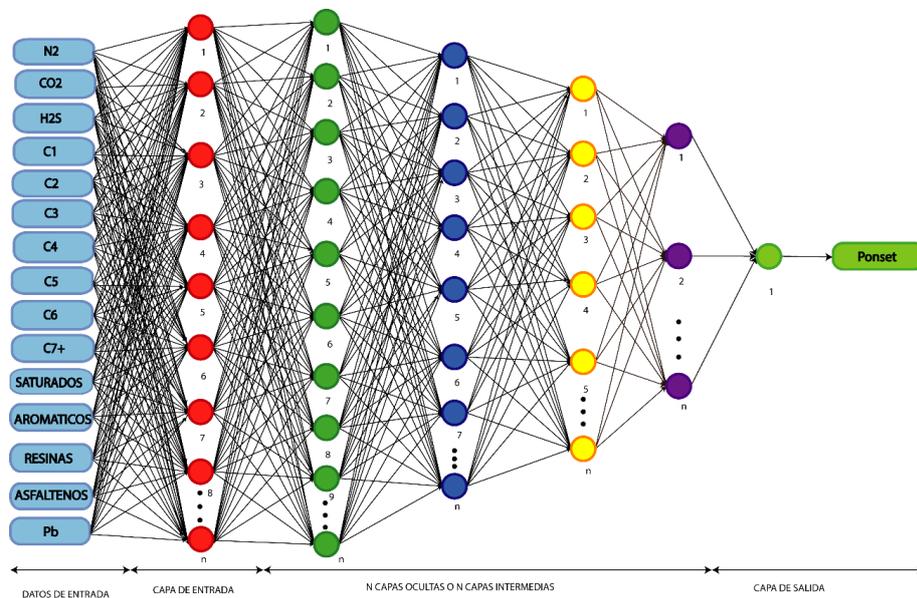


Figura 3.9 Arquitectura de la red neuronal AsphPy.

Para la construcción de la red neuronal, se declaró cuantas neuronas o grafos tendría la capa de entrada, la cantidad de capas ocultas y por último se declaró la capa de salida, esto es posible mediante el método o función "Sequential()".

Para cada capa se utilizaron dos funciones, la primera función es “*add*” y “*Dense()*”, para comprender mejor cómo realizar o implementar la arquitectura full connected se ilustra en la Figura 3.11 un fragmento del código.

```
model = Sequential()
model.add(Dense(50, input_dim=16, activation='relu'))
model.add(Dense(45, input_dim=50, activation='relu'))
model.add(Dense(40, input_dim=45, activation='relu'))
model.add(Dense(35, input_dim=40, activation='relu'))
model.add(Dense(30, input_dim=35, activation='relu'))
model.add(Dense(25, input_dim=30, activation='relu'))
model.add(Dense(20, input_dim=25, activation='relu'))
model.add(Dense(15, input_dim=20, activation='relu'))
model.add(Dense(1, activation='relu'))
```

Figura 3.10 Declaración de las capas en la red neuronal.

Inicialmente se creó el modelo, a dicho modelo se fueron agregando las capas mediante la función “*add*” y gracias a “*Dense()*” se les declaro los respectivos parámetros a cada capa como: número de neuronas, numero de la dimensión de entrada a la capa y como tercer parámetro se definió que función de activación.

La red AsphPy cuenta con una capa de entrada con 50 neuronas y una dimensión de entrada igual a 16 que corresponden al número de variables predictoras, lo que compete a las capas ocultas, AsphPy cuenta con 7 capas ocultas para cada capa se le definió los parámetros mencionados anteriormente y por último se definió la capa de salida que cuenta con una neurona.

3.3.5. Función de perdidas, optimizador y métricas.

Una vez declarada la arquitectura de la red, se procedió a definir la función de perdías, la función de optimización y que métrica se usó al momento de realizar el entrenamiento de la red neuronal. Dos de los tres parámetros para la creación del modelo de la red neuronal son modelos o funciones de estadística que evaluarán el rendimiento de la red al momento de comparar los datos predichos y los datos reales que se utilizaron para la predicción. Para esto es necesario describir cada parámetro

Función de perdidas.

La función de perdida es un método para la evaluación del rendimiento de un algoritmo de predicción. Este estadístico evalúa que tan preciso son los datos de predicción con los datos reales que se utilizaron para la elaboración del modelo

del algoritmo. Existen varias funciones de pérdida y cada una depende de la aplicación en la que se va a implementar el algoritmo de predicción. Para AsphPy se implementó una función de pérdidas llamada error porcentual absoluto medio (mean absolute percentage error) y es ilustrado en la Ecuación 3.2.

```
model.compile(loss='mean_absolute_percentage_error',  
              optimizer='adam',  
              metrics='mae')
```

Figura 3.11 Función de pérdidas, optimizador y métrica de evaluación para la creación del modelo.

$$E = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad \text{Ecuación 3.2}$$

Optimizador.

El optimizador realiza la maximización de la función de pérdidas, es decir, reduce las pérdidas en el modelo de predicción. Igual que la función de pérdidas, existen una gran cantidad de funciones y cada una depende de la aplicación en la que se desea implementar el algoritmo. En AsphPy se implementó un optimizador llamado Adam, este optimizador adapta un radio de aprendizaje en función de los patrones de distribución de los parámetros, es decir, si los parámetros o características en los datos están muy dispersos, el radio de aprendizaje aumentara. Adam es un optimizador de aprendizaje adaptativo.

Métricas.

Las métricas son funciones que permiten evaluar de forma general el rendimiento del algoritmo de predicción, Son funciones que tienen como finalidad calcular la precisión de la red neuronal y que la misma realice su modelo bajo estos parámetros de rendimiento. Existen dos métricas muy usadas en el mundo de la ciencia de datos los cuales son; el accuracy, este parámetro solo tomara de 0 a 100%, siendo 100% una excelente precisión; otra métrica o función estadística muy importante es la media del error absoluto, como su nombre nos indica ayuda a calcular el error promedio que existe entre un set de datos de entrenamiento y los datos de predicción. En AsphPy se utilizó la segunda métrica mencionada anteriormente y es ilustrado en la Ecuación 3.3.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad \text{Ecuación 3.3}$$

3.3.6. Entrenamiento de la red neuronal.

Una vez se realizó el modelo de la red neuronal y se manipuló la base de datos, se entrenó la red neuronal con el conjunto de datos determinado previamente. Para el entrenamiento de una red neuronal es muy importante el parámetro épocas, este parámetro indica cada iteración que se hace para encontrar la relación entre los datos de entrada y los de salida. Para observar el rendimiento de este entrenamiento se toma en cuenta los valores que toma la función de pérdidas a medida que se hacían las iteraciones. En la Figura 3.12 se observa el comportamiento de dicho parámetro.

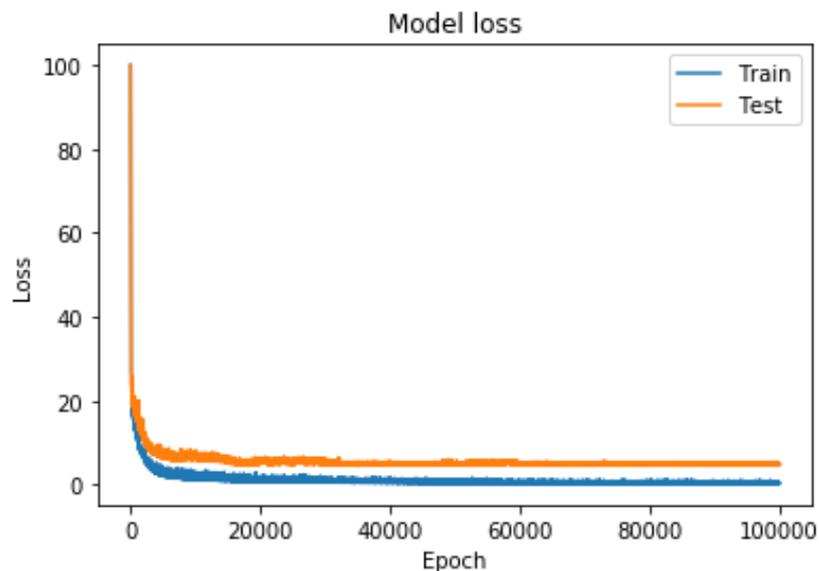


Figura 3.12 Representación gráfica de las pérdidas en el modelo de predicción.

En la Figura 3.12, se puede observar como la red neuronal tiende a aprender los rasgos más generales que se encuentran en dichos datos reflejado con la disminución de los valores calculados con la función de pérdida.

3.4. MÁQUINAS DE SOPORTE VECTORIAL.

3.4.1. Datos para la formulación del modelo.

Para evaluar el rendimiento entre la red neuronal AsphPy y las máquinas de soporte vectorial, se partió de los mismos datos utilizados en el entrenamiento de la red. Esto tiene como finalidad evaluar los dos algoritmos en las mismas condiciones posibles. Así poder determinar que algoritmo tiene mejor rendimiento.

Para esto se utilizó los identificadores del subconjunto de entrenamiento declarados anteriormente. Previo a esto, se realizó una nueva base de datos igual a la que se utilizó en la red neuronal; esto tuvo como finalidad tener los datos de presiones en unidades de MPa visto que al momento de implementar el SVM en unidades de psi, las predicciones arrojaban errores significativamente alarmantes, una vez se predecían los datos de interés, se realizó una conversión de unidades para tener las presiones reales, las de AsphPy y las de SVM en las mismas unidades y así poder comparar los datos de presión Onset.

A continuación, en la Figura 3.13 se ilustra la primera fracción de código que compete solo a la máquina de soporte vectorial. En dicho fragmento se podrá observar cómo se importan los datos de Excel a Python y como se declaran las variables predictoras y de interés para la realización del modelo.

```
predictoras1 = pd.read_excel("../datasets/predictorasint.xlsx")
ponset1 = pd.read_excel("../datasets/ponsetsint.xlsx")
Muestras1 = pd.concat([predictoras1,ponset1], axis = 1)
Muestras1.head()
X_vals1 = np.array(predictoras1, "float32")
Y_vals1 = np.array(Muestras1["Ponset"],"float32")
x_vals_train1 = X_vals1[train_id]
y_vals_train1 = Y_vals1[train_id]
```

Figura 3.13 Lectura de la base de datos en Excel y declaración de las variables de entrenamiento.

Las variables “*x_vals_train1*” y “*y_vals_train1*” contiene las mismas muestras de crudo utilizadas en el entrenamiento de la red para la creación de su modelo, esto tiene como fin crear los dos modelos de predicción con los mismos datos utilizados. Para garantizar que sean iguales las mismas muestras, se utilizaron los mismos identificadores almacenados en la variable “*train_id*”.

3.4.2. Kernel y Modelo de predicción SVM.

Para la creación del modelo de predicción en machine learning se implementó la función de base radial (rbf). Dicho kernel realizó el hiperplano en el cual albergo los datos o muestras de crudo que se utilizaron para la creación de AsphPy. La Ecuación 3.4 ilustra la función del kernel utilizado en este algoritmo.

$$K_{(x,x')} = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$$

Ecuación 3.4

Donde $\|x - x'\|^2$ es la distancia euclidiana al cuadrado de los dos vectores implementados en el hiperplano y σ es un parámetro libre. A la hora de implementar el modelo en Python de las máquinas de soporte vectorial es necesario declarar la constante C y los parámetros necesarios para la realización del modelo esto es ilustrado en la Figura 3.16.

```
C = 1e3  
svr_rbf = SVR(kernel="rbf", C=C, gamma=0.1)
```

Figura 3.14 Declaración de parámetros para las máquinas de soporte vectorial.

Donde C es una constante que debe ser mayor a 0 y determina el equilibrio entre la regularidad de la función implementada para la creación del hiperplano y a los valores de las desviaciones mayores que la banda de soporte y gamma la constante definida anteriormente.

Posteriormente se definió el modelo de predicción en Python ilustrado en la Figura 3.15.

```
model_rbf = svr_rbf.fit(x_vals_train1,y_vals_train1)
```

Figura 3.15 Declaración de parámetros para las máquinas de soporte vectorial.

La función para la creación del modelo tiene como parámetros las muestras que se utilizaron en el entrenamiento de la red neuronal “x_vals_train1” y “y_vals_train1”.

4. RESULTADOS Y DISCUSIONES

4.1. VALIDACIÓN DEL MODELO

Para evaluar el rendimiento de los modelos de predicción que se realizaron en el proyecto, es decir, el modelo de la red neuronal AsphPy y el algoritmo de machine learning implementado, se calculó el porcentaje de error entre la presión onset real y las predicciones de los modelos. El porcentaje de error está definido por la Ecuación 4.1.

$$\text{Error (\%)} = \left[\frac{\text{Presión onset real} - \text{Presión onset predicha}}{\text{Presión onset real}} \right] \times 100 \quad \text{Ecuación 4.1}$$

Se escogieron 100 muestras para ilustrar el comportamiento de esta métrica de evaluación. En el anexo, también se ilustra tabla con los errores calculados entre la red neuronal y el SVM con respecto a la presión onset real de las 100 muestras escogidas aleatoriamente.

Una vez se calculó el error existente entre la presión real y la predicción en cada una de las muestras de crudo, se calculó el promedio de los errores en las 345 muestras para observar o calcular el rendimiento general de los modelos.

Tabla 4.1 Error promedio de los dos modelos de predicción.

	Red Neuronal AsphPy	Máquinas de soporte vectorial
Error promedio (%)	2.00	12.61703

Como se observó en la tabla 4.3, la red neuronal AsphPy presentó un mejor rendimiento a la hora de predecir la presión onset partiendo de ciertas variables de entrada en comparación con el algoritmo de machine learning implementado, en este caso las máquinas de soporte vectorial.

Una vez se calculó el rendimiento de los modelos de predicción se escogió 8 muestras de crudo para observar el comportamiento que presentó la presión onset a medida que la temperatura en el yacimiento aumenta. Se definieron los mismos rangos de temperatura que se utilizaron para la creación de los modelos de predicción, es decir, 320K, 340K, 360K, 380K y 400K. En las Figuras 4.1 a la Figura 4.8, se puede observar las tres presiones onset (upper); la presión real,

la presión predicha por la red neuronal y la presión predicha por las máquinas de soporte vectorial, al mismo tiempo la Tabla 4.2 y Tabla 4.3, se encuentra la composición química y el análisis SARA junto a la presión de saturación de las 8 muestras.

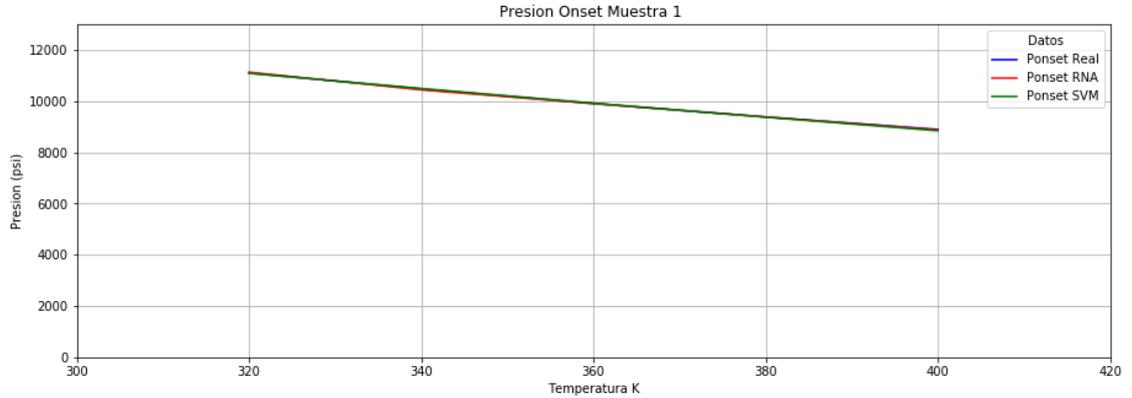


Figura 4.1 Presión vs temperatura muestra (1) presión Onset (upper).

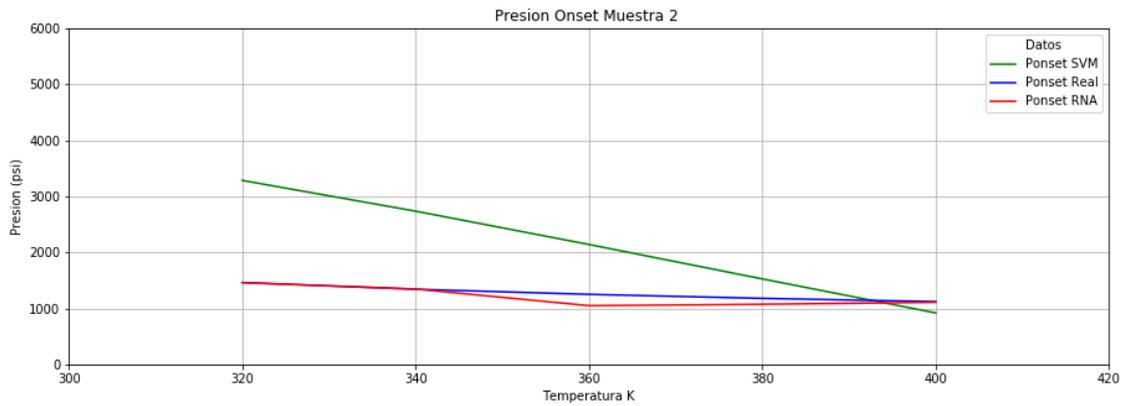


Figura 4.2 Presión vs temperatura muestra (2) presión Onset (upper).

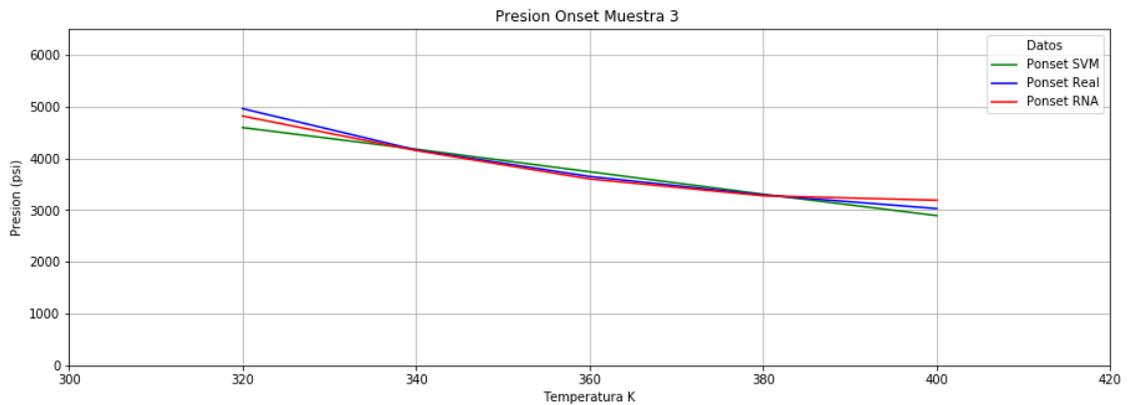


Figura 4.3 Presión vs temperatura muestra (3) presión Onset (upper).

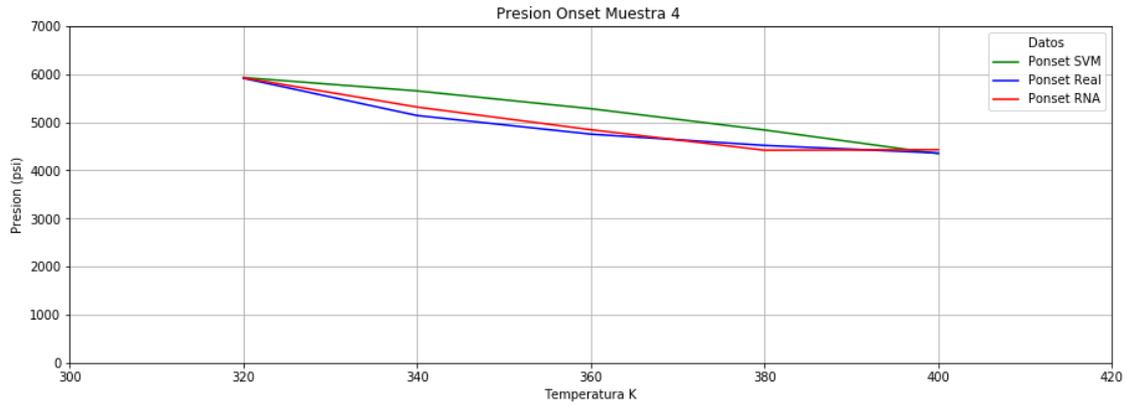


Figura 4.4 Presión vs temperatura muestra (4) presión Onset (upper).

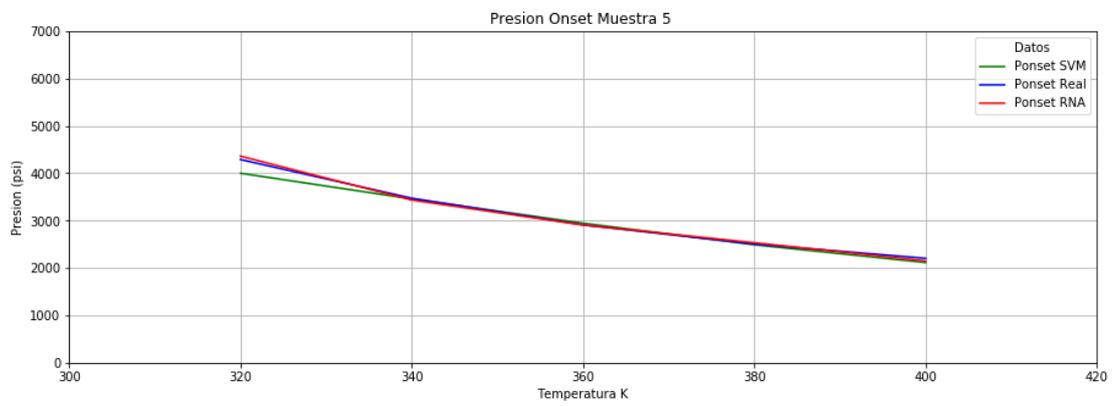


Figura 4.5 Presión vs temperatura muestra (5) presión Onset (upper).

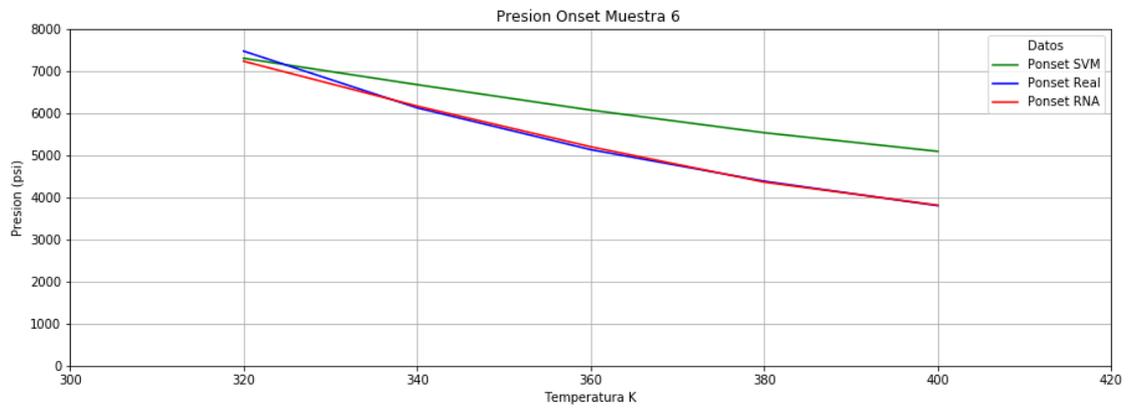


Figura 4.6 Presión vs temperatura muestra (6) presión Onset (upper).

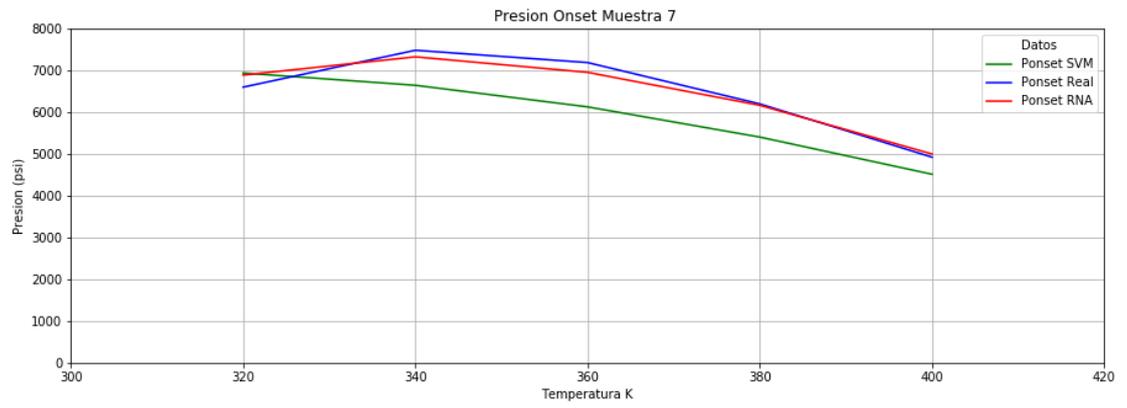


Figura 4.7 Presión vs temperatura muestra (7) presión Onset (upper).

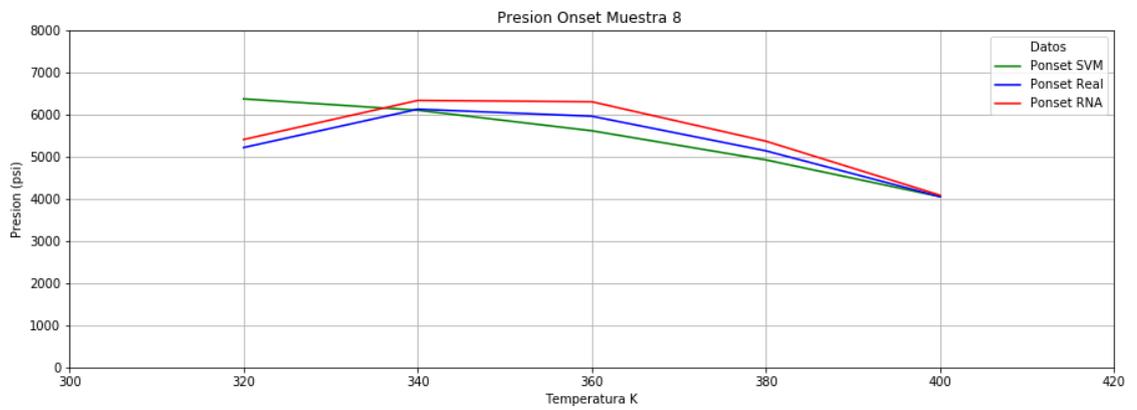


Figura 4.8 Presión vs temperatura muestra (8) presión Onset (upper).

Tabla 4.2 Composición química de las 8 muestras escogidas aleatoriamente.

N°	T	N2	CO2	H2S	C1	C2	C3	C4	C5	C6	C7Plus
1	320	1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
1	340	1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
1	360	1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
1	380	1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
1	400	1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
2	320	0.10	10.6	0	11.2887	5.0841	6.1839	3.627	7.965	3.105	51.993
2	340	0.10	10.6	0	11.2887	5.0841	6.1839	3.627	7.965	3.105	51.993
2	360	0.10	10.6	0	11.2887	5.0841	6.1839	3.627	7.965	3.105	51.993
2	380	0.10	10.6	0	11.2887	5.0841	6.1839	3.627	7.965	3.105	51.993
2	400	0.10	10.6	0	11.2887	5.0841	6.1839	3.627	7.965	3.105	51.993
3	320	0.08	30.5	0	8.7801	3.9543	4.8097	2.821	6.195	2.415	40.439
3	340	0.08	30.5	0	8.7801	3.9543	4.8097	2.821	6.195	2.415	40.439
3	360	0.08	30.5	0	8.7801	3.9543	4.8097	2.821	6.195	2.415	40.439
3	380	0.08	30.5	0	8.7801	3.9543	4.8097	2.821	6.195	2.415	40.439
3	400	0.08	30.5	0	8.7801	3.9543	4.8097	2.821	6.195	2.415	40.439
4	320	0.36	0.17	0	35.733	8.8519	6.8208	1.227	3.801	8.037	34.984
4	340	0.36	0.17	0	35.733	8.8519	6.8208	1.227	3.801	8.037	34.984
4	360	0.36	0.17	0	35.733	8.8519	6.8208	1.227	3.801	8.037	34.984
4	380	0.36	0.17	0	35.733	8.8519	6.8208	1.227	3.801	8.037	34.984
4	400	0.36	0.17	0	35.733	8.8519	6.8208	1.227	3.801	8.037	34.984
5	320	0.16	1.94	0	33.6	7.557	6.742	6.579	5.179	5.049	32.631
5	340	0.16	1.94	0	33.6	7.557	6.742	6.579	5.179	5.049	32.631
5	360	0.16	1.94	0	33.6	7.557	6.742	6.579	5.179	5.049	32.631
5	380	0.16	1.94	0	33.6	7.557	6.742	6.579	5.179	5.049	32.631
5	400	0.16	1.94	0	33.6	7.557	6.742	6.579	5.179	5.049	32.631
6	320	0.08	13.3	4.743	41.256	4.815	3.789	3.573	2.871	3.204	22.14
6	340	0.08	13.3	4.743	41.256	4.815	3.789	3.573	2.871	3.204	22.14
6	360	0.08	13.3	4.743	41.256	4.815	3.789	3.573	2.871	3.204	22.14
6	380	0.08	13.3	4.743	41.256	4.815	3.789	3.573	2.871	3.204	22.14
6	400	0.08	13.3	4.743	41.256	4.815	3.789	3.573	2.871	3.204	22.14
7	320	0.44	1.97	0.000	23.333	8.301	6.049	5.036	7.520	5.176	42.167
7	340	0.44	1.97	0.000	23.333	8.301	6.049	5.036	7.520	5.176	42.167
7	360	0.41	1.85	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429
7	380	0.44	1.97	0.000	23.333	8.301	6.049	5.036	7.520	5.176	42.167
7	400	0.44	1.97	0.000	23.333	8.301	6.049	5.036	7.520	5.176	42.167
8	320	0.41	1.85	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429
8	340	0.41	1.85	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429
8	360	0.41	1.85	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429
8	380	0.41	1.85	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429

Tabla 4.3 Análisis SARA y presión de saturación (Pb) de las 8 muestras escogidas aleatoriamente.

SATURADOS	AROMATICOS	RESINAS	ASFALTENOS	Pb
59.30	37	2.7	0.9	748.873051
59.30	37	2.7	0.9	873.221443
59.30	37	2.7	0.9	1037.93438
59.30	37	2.7	0.9	1259.02594
59.30	37	2.7	0.9	1560.30211
38.80	49.6	7.7	3.9	574.395369
38.80	49.6	7.7	3.9	732.876819
38.80	49.6	7.7	3.9	869.658693
38.80	49.6	7.7	3.9	988.684168
38.80	49.6	7.7	3.9	1093.02071
38.80	49.6	7.7	3.9	1282.55305
38.80	49.6	7.7	3.9	1426.29052
38.80	49.6	7.7	3.9	1538.48839
38.80	49.6	7.7	3.9	1628.56243
38.80	49.6	7.7	3.9	1702.39217
39.20	35.9	9	15.5	1924.65084
39.20	35.9	9	15.5	2129.50357
39.20	35.9	9	15.5	2312.42076
39.20	35.9	9	15.5	2478.5303
39.20	35.9	9	15.5	2631.2379
66.30	25.6	5.4	2.8	1685.47793
66.30	25.6	5.4	2.8	1796.91957
66.30	25.6	5.4	2.8	1887.82066
66.30	25.6	5.4	2.8	1963.50005
66.30	25.6	5.4	2.8	2027.51777
73.40	19.3	7.1	0.2	3084.53574
73.40	19.3	7.1	0.2	3416.49025
73.40	19.3	7.1	0.2	3678.98994
73.40	19.3	7.1	0.2	3893.03726
73.40	19.3	7.1	0.2	4071.42381
24.8	45.6	16.8	12.8	1416.4503
24.8	45.6	16.8	12.8	1568.1026
24.8	45.6	16.8	12.8	1536.5193
24.8	45.6	16.8	12.8	1799.7166
24.8	45.6	16.8	12.8	1889.7230
24.8	45.6	16.8	12.8	1287.6840
24.8	45.6	16.8	12.8	1428.7371
24.8	45.6	16.8	12.8	1536.5193
24.8	45.6	16.8	12.8	1643.1330
24.8	45.6	16.8	12.8	1720.0774

Para observar las presiones de todas las 345 muestras se organizaron de menor a mayor presión y se procedió a graficar, con la finalidad de poder observar las tendencias de las predicciones de los dos modelos con respecto a la presión real en el yacimiento. En la Figura 4.9. se compara las presión real con respecto a las predicciones por la red neuronal AsphPy; la Figura 4.10. contiene la presión onset real y la predicción por las máquinas de soporte vectorial y finalmente en la Figura 4.11. se observa la presión onset real y las presiones predichas por los dos modelos realizados en este trabajo.

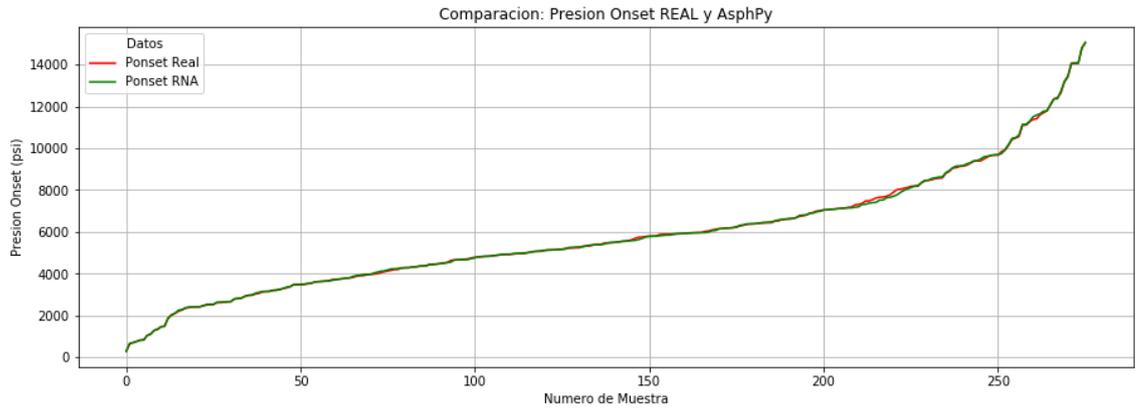


Figura 4.9 Presión real vs predicciones de la red neuronal.

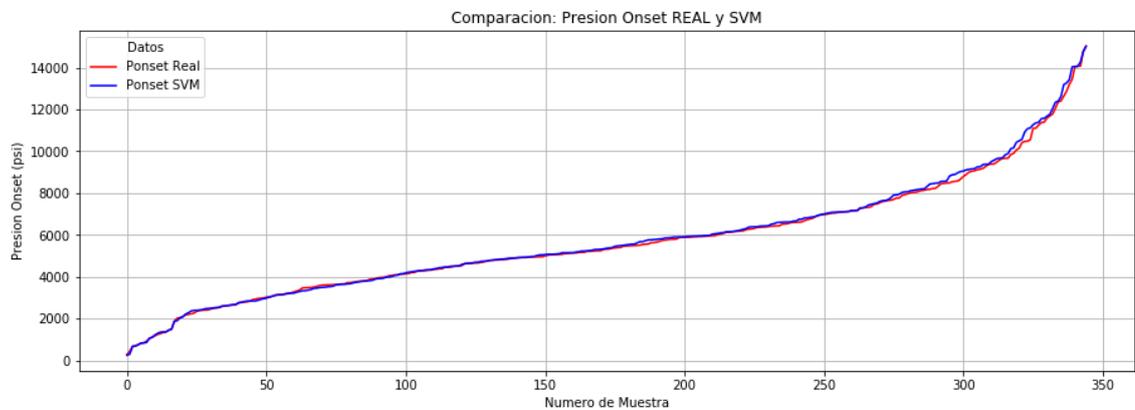


Figura 4.10 Presión real vs predicciones de las máquinas de soporte vectorial.

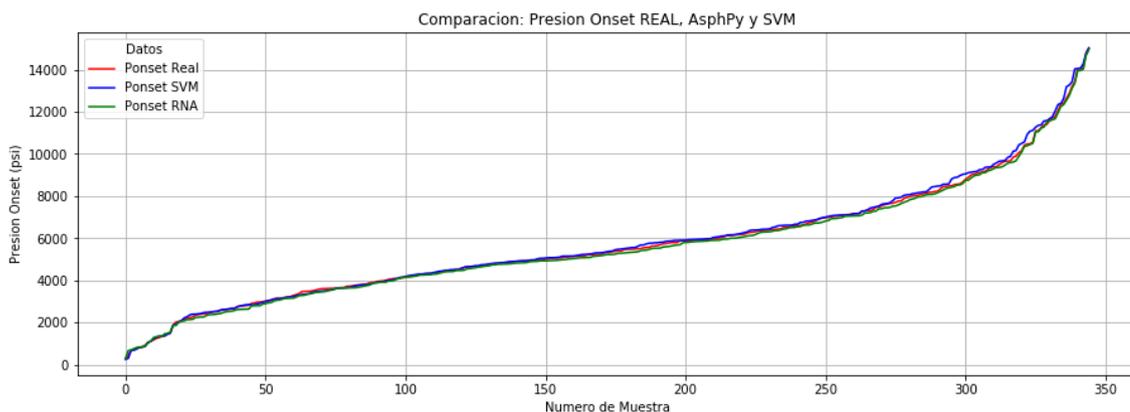


Figura 4.11 Presión real vs predicciones de los modelos de predicción.

Al momento que se realizó o se definió los parámetros para la elaboración del modelo de la red neuronal se varió uno de los parámetros más importantes a la hora de elaborar y construir el modelo de predicción en la red neuronal; dicho parámetro es la cantidad de épocas o cantidad de iteraciones que se realizaron en la etapa de entrenamiento. Esto tuvo como finalidad observar la fiabilidad o exactitud de la red neuronal a la hora de predecir la presión onset una vez entrenada dicha red. Como se puede observar en la Tabla 4.4 este parámetro juega un papel importante, visto que si la cantidad de épocas es un valor pequeño el error máximo o promedio será muy elevado, lo que nos dice que la red neuronal no será tan exacta a la hora de hacer la predicción; por el contrario, si se eleva este número de iteraciones, la red neuronal tiende a converger y mejorar las predicciones con respecto a los datos reales. Si el número es muy elevado sucede algo muy particular; la red neuronal empieza a perder convergencia y el error aumentara, a esto se le conoce como sobre entrenamiento, la red neuronal solo aprenderá las características o patrones de solo los datos que se utilizaron en esta etapa y al momento de validar con datos nuevos datos la red tendrá predicciones erradas. Por tal razón el número de iteraciones no puede ser muy bajo o muy elevado.

Tabla 4.4 Error máximo y error promedio con respecto al número de épocas para entrenamiento de la red neuronal.

Épocas	Error máximo	Error promedio	Cantidad de muestras por encima del 5% de error
500	86.44%	13.90%	202
1000	82.17%	10.20%	172
5000	63%	5.50%	95
10000	64.60%	3.20%	38
50000	83.90%	2.70%	36
100000	42.30%	2.00%	31
200000	Mayor a 100%	2.4%	34

4.2. ANÁLISIS DEL COMPORTAMIENTO DE LAS MUESTRAS CUANDO SE INYECTA GAS.

La tabla 4.4 ilustra el error calculado entre las muestras que se simularon en CMG y los datos experimentales. Para este caso, del total de muestras que se adicionaron en este estudio, se escogieron 15 muestras partiendo de la muestra inicial de yacimiento, la muestra al inyectarle 10% CO₂ y 5% gas asociado.

Del conjunto de simulaciones se observa que, a altas temperaturas la ecuación de Peng Robinson (1978) no tiene buenas aproximaciones cuando se hace el ajuste de las condiciones liquido-vapor, por este motivo se decide usar Peng Robinson (1976). Además de esto, cuando hay algún tipo de inyección de gas al yacimiento, la calibración es mucho más compleja incluso si el contenido de CO₂ aumenta, hay un error bastante alto antes de iniciar el modelamiento de los asfaltenos ajustando las fases solido-liquido.

Esto se debe principalmente a que la ecuación de PR-1976 es altamente usada en la industria del petróleo en la simulación de sistemas de gas, cumpliendo con varios objetivos, comparándola con la SRK. Los parámetros son expresados en función de las propiedades críticas y el factor acéntrico, el modelo es razonable a condiciones cercanas al punto crítico y las reglas de mezclas deben usar un solo parámetro sobre las interacciones binarias, las cuales no tenían que ser dependiente de la presión, temperatura y composición.

Tabla 4.5. Cálculo del porcentaje de error en las muestras simuladas en CMG

Tipo de muestra		T	Pb (psi)	AOP (psi)	CMG WinProp		% error Pb	% error AOP
					Pb	AOP		
Inicial		320	1173.05	4127.35	1215.00	4214.70	3.576	2.116
		340	1297.92	5033.24	1267.00	5017.00	2.382	0.323
		360	1402.81	4958.09	1412.00	5015.00	0.655	1.148
		380	1493.29	4262.28	1613.00	4413.00	8.016	3.536
		400	1574.26	3323.82	1617.00	3413.00	2.715	2.683
CO ₂	10%	320	1272.24	3906.17	1215.00	4012.00	4.499	2.709
		340	1419.35	4835.69	1412.00	5008.00	0.518	3.563
		360	1577.52	4945.68	1615.00	5013.00	2.376	1.361
		380	1701.92	4454.37	1811.00	4607.00	6.409	3.427
		400	1806.62	3722.50	1820.00	3809.00	0.740	2.324
Gas Asociado	5%	320	1287.68	5217.51	1213.00	5215.00	5.800	0.048
		340	1428.74	6127.05	1511.00	6507.00	5.758	6.201
		360	1536.52	5960.50	1619.00	6011.00	5.368	0.847
		380	1643.13	5137.96	1607.00	5213.00	2.199	1.461
		400	1720.08	4051.85	1814.00	4209.00	5.460	3.879

Según Bahrami (Peyman Bahrami, 2015), la presión Onset de una muestra aumenta a cualquier temperatura cuando algún gas (C_1 , N_2 , gas asociado, etc.) es agregado a la mezcla. Pero este comportamiento no ocurre de igual forma con el CO_2 : a temperaturas por encima del "crossover point" este gas actúa como un inhibidor y la curva de la envolvente de la presión onset tendrá pendiente negativa, o en su defecto igual a 0. Por otra parte, si la temperatura está por encima de determinado punto, actúa como un agente precipitante muy fuerte con una pendiente de la curva positiva respectivamente. La variación de esta tendencia en las curvas de precipitación de la Presión onset, es el comportamiento más destacado de este gas.

Comparando el mismo %molar del N_2 y CO_2 , es evidente que el N_2 tiene su envolvente cubriendo una parte más alta de la gráfica (incluso a bajas concentraciones) a temperaturas bajas. Pero el CO_2 a temperaturas altas es un agente precipitante que conduce a un mayor contenido de asfaltenos para depositar.

4.3. ANÁLISIS DE LAS TENDENCIAS DE LAS CURVAS AOP

Otro aspecto importante es la tendencia de las curvas que forman los cinco puntos de la curva de presión Onset evaluados a distintas temperaturas. Anteriormente se tenían muestras con una tendencia descendente, es decir, que la presión onset de todas las muestras disminuía con el aumento de la temperatura. Es allí donde se decide buscar nuevas tendencias de crudos que se comporten de otra manera con la finalidad de conformar una red neuronal más rigurosa, estas muestras agregadas tienen una tendencia de envolvente. A continuación, se mostrarán los cuatro grupos que fueron encontrados para una mejor caracterización de las muestras, para cada grupo por medio de Python fueron graficadas las distintas variables de datos de entrada y finalmente se obtuvieron distintas deducciones respectivamente:

4.3.1 Grupo 1

La Figura 4.12 corresponde al grupo 1, en donde las curvas forman una envolvente, es decir, que a bajas temperaturas la Presión onset aumenta, llega a un pico máximo y vuelve a caer a temperaturas altas.

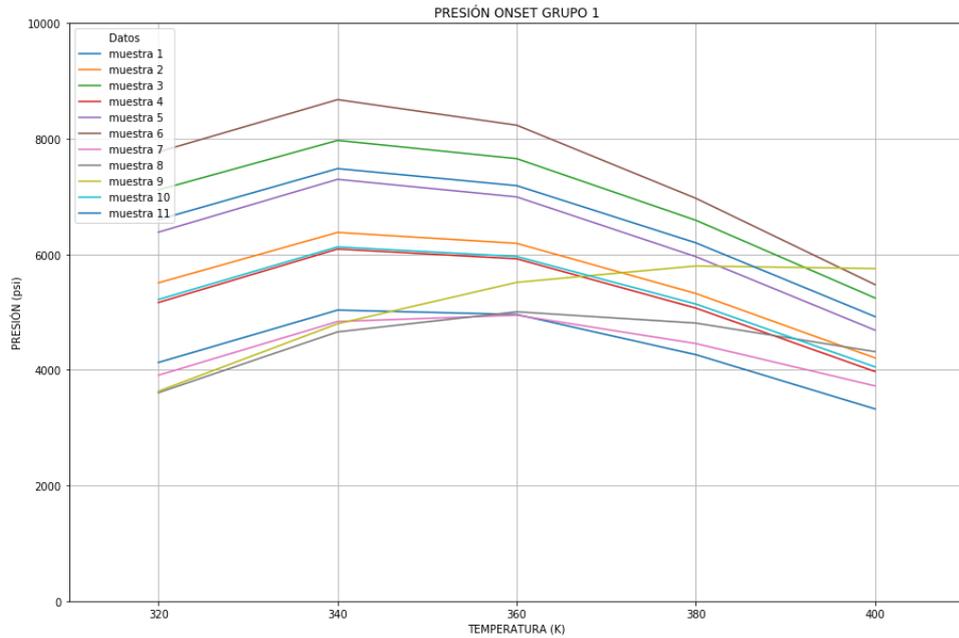


Figura 4.12 Gráfica AOP vs Temperatura Grupo 1.

Según la Figura 4.13, las muestras de este grupo tienen una presión de burbuja por encima de 1100 psi y por debajo de 3000 aproximadamente, con una pendiente leve.

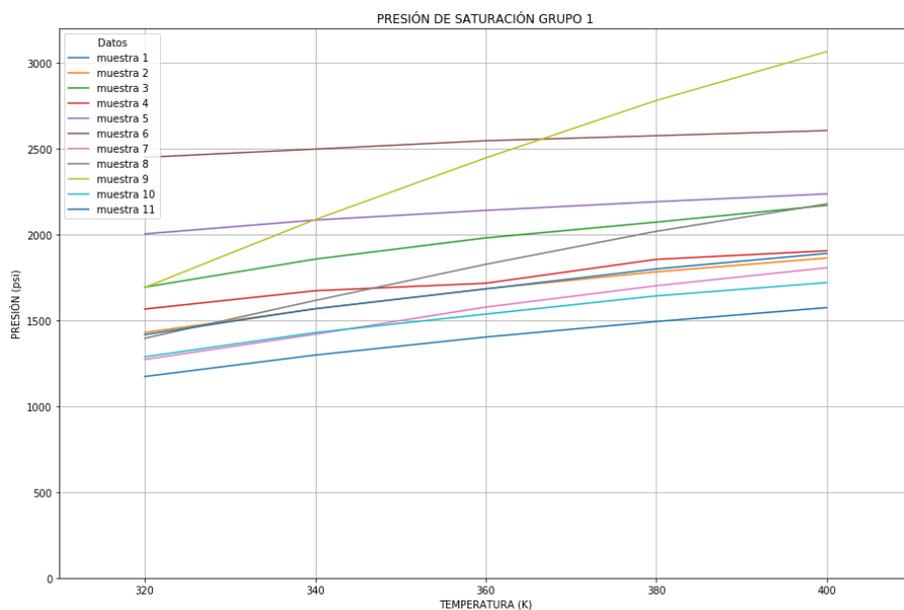


Figura 4.13 Tendencia de la presión de saturación del grupo 1 con respecto a la variación de temperatura.

En la Figura 4.14 se evidencia que todas las muestras presentan el mismo análisis Sara, debido a que al mismo crudo se le evaluó la incidencia al inyectarle gas de distinta composición y proporción al yacimiento, según el estudio de (Peyman Bahrami, 2015). Además de eso se puede observar que en el análisis SARA, el contenido de aromáticos es casi del 50%. Y teniendo en cuenta el

índice de estabilidad coloidal (CII), estas muestras son estables pues tienen un $CII < 0.7$ a condiciones de yacimiento; pero cuando ya se alteran propiedades como presión y temperatura, ya comienzan a aparecer problemas relacionados con las depositación.

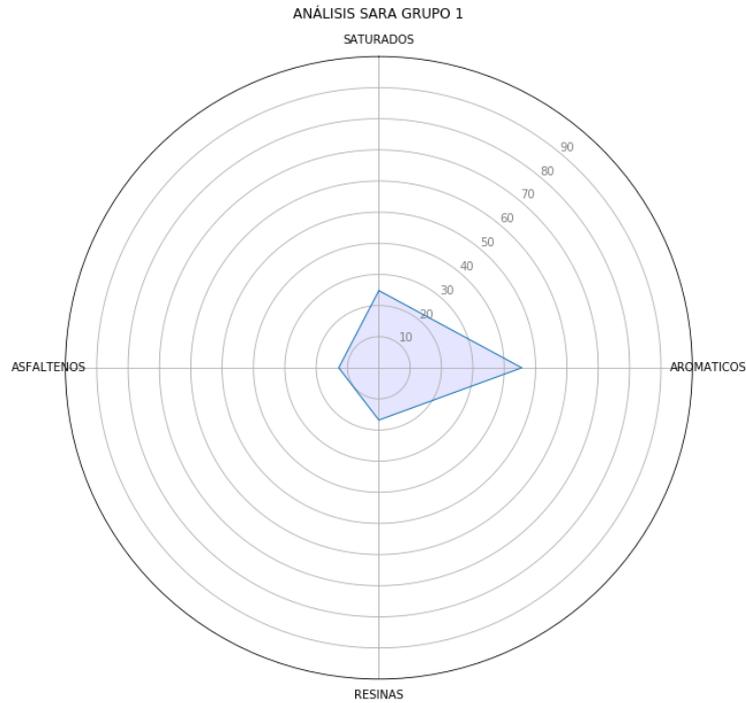


Figura 4.14 Análisis SARA del grupo 1.

La Figura 4.15 muestra una tendencia definida en donde el %molar del C1 está entre en el rango de (18-20) %, los componentes no hidrocarburos están por debajo del 10% y el porcentaje de C7+ por encima del 40%.

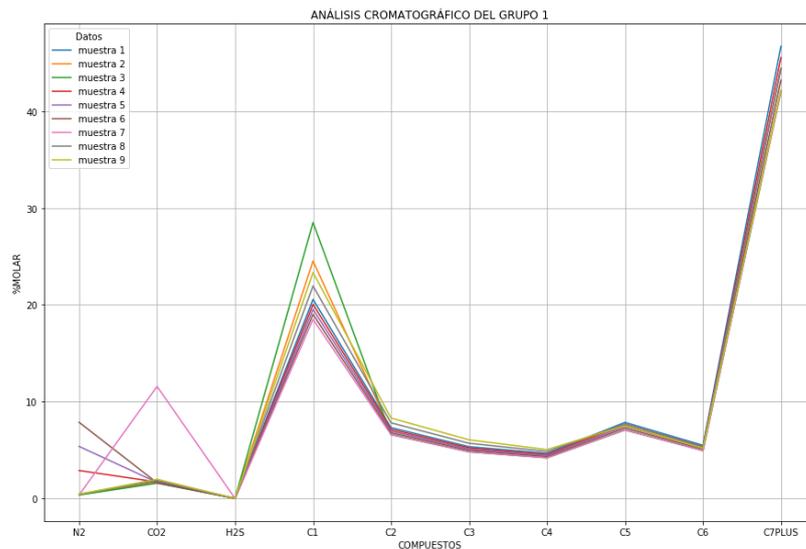


Figura 4.15 Análisis cromatográfico del grupo 1.

4.3.2 Grupo 2

El segundo grupo de curvas pertenece a líneas que tienen una tendencia plana o su pendiente es muy baja en el rango de temperatura establecido, ver Figura 4.16.

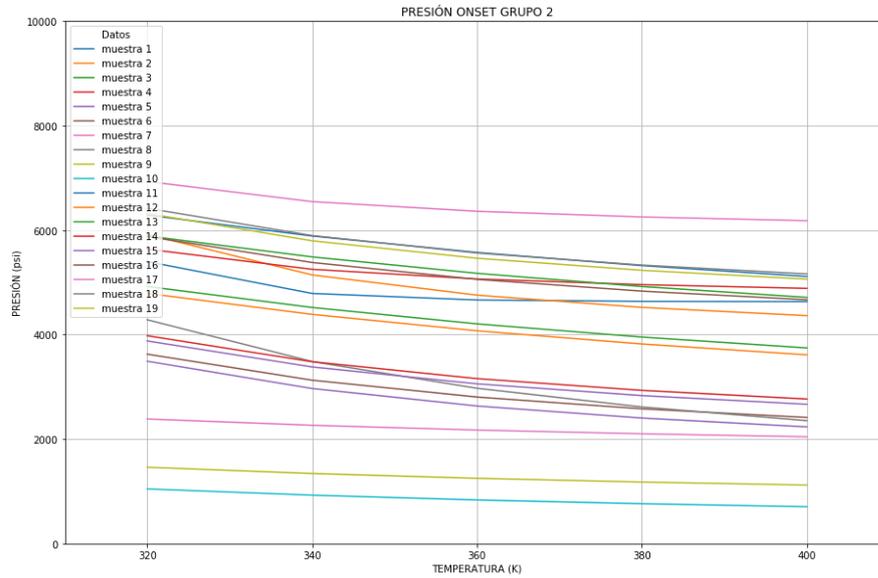


Figura 4.16 Gráfica AOP vs Temperatura Grupo 2.

En cuanto a la presión de burbuja, en la Figura 4.17 se observa que para el grupo 2 la pendiente es suave parecida a la del grupo anterior, pero con la diferencia de que ya las presiones ya son más elevadas, llegando incluso a 4100 psi.

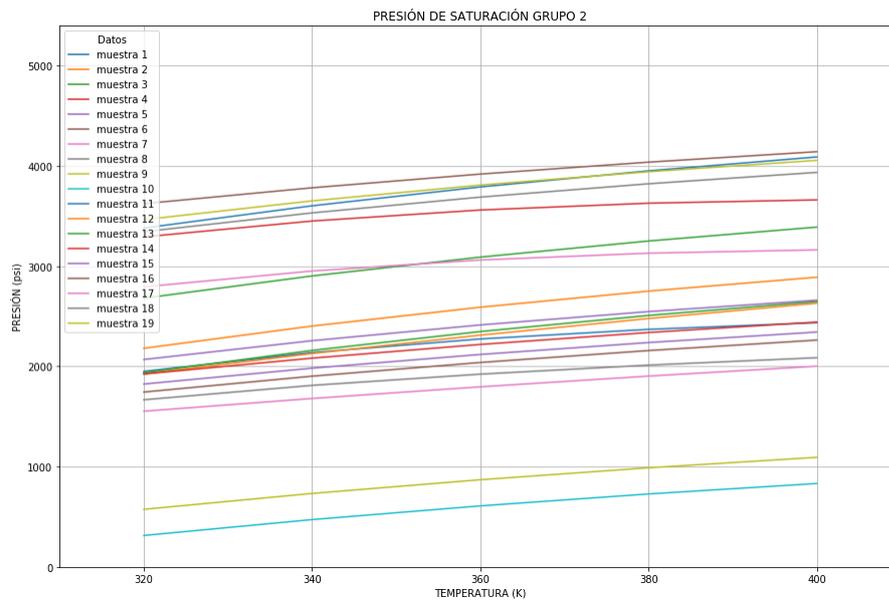


Figura 4.17 Tendencia de la presión de saturación del grupo 2 con respecto a la variación de temperatura.

En la Figura 4.18 se evidencia que hay una fuerte tendencia de los saturados presentes en las muestras del grupo 3, algunas con un porcentaje mayor al 50%. Otra característica importante es el poco porcentaje de resinas encargadas de mantener en suspensión los asfaltenos y separados entre ellos mismos, debido a que estas forman una capa alrededor de los asfaltenos que genera una fuerza de repulsión mayor a la fuerza de atracción. Por tal motivo se observa en la gráfica de la Figura 4.16 que a bajas presiones ya hay precipitación de este compuesto, pues las resinas no alcanzan a cubrir la totalidad de los asfaltenos.

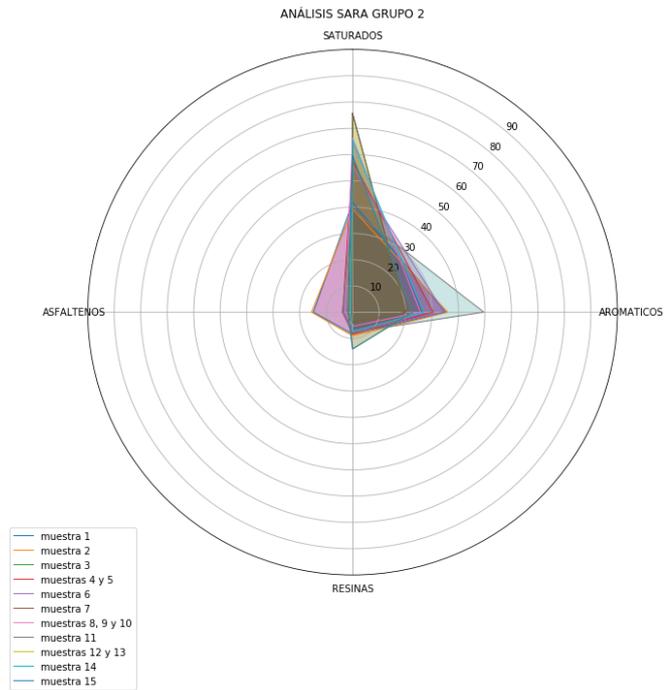


Figura 4.18 Análisis SARA del grupo 2.

En la Figura 4.19, se muestra que el porcentaje de Metano en las muestras es alto, encontrándose en un rango de (27-47) %, C7+ por debajo del 40%.

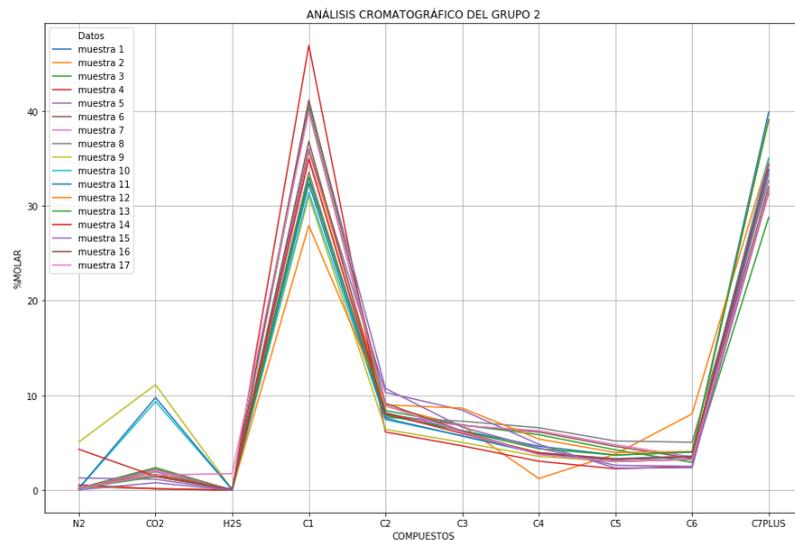


Figura 4.19 Análisis cromatográfico del grupo 2.

4.3.3. Grupo 3

El tercer grupo de curvas se caracteriza por tener una pendiente negativa más pronunciada que los dos grupos anteriormente mostrados, la inclinación es un poco más pronunciada. Además de eso hay algunas presiones onset altas, de hasta 14000 psi, tal y como se evidencia en la Figura 4.20.

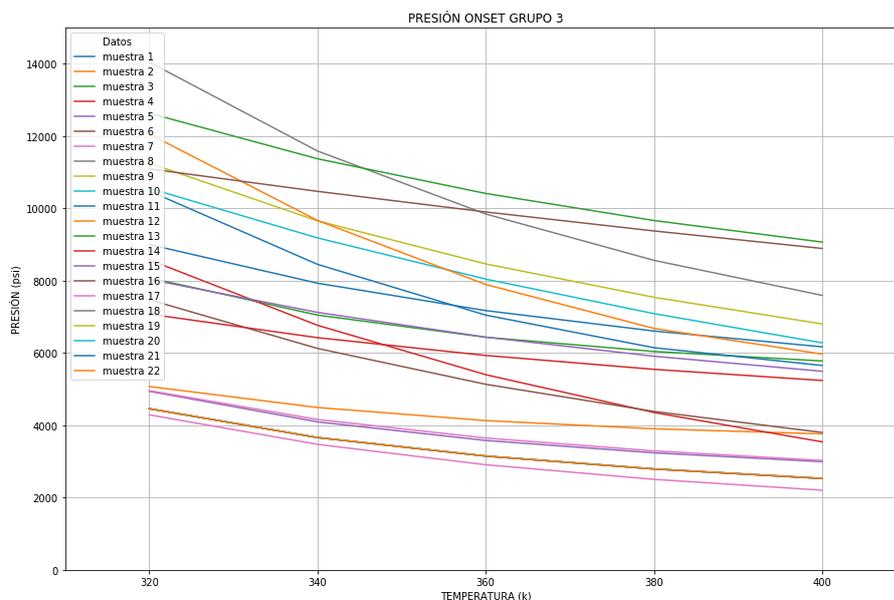


Figura 4.20 Gráfica AOP Vs Temperatura Grupo 3.

Por otra parte, la presión de burbuja presenta una tendencia parecida entre las mismas muestras pero sin un rango definido, este comportamiento se puede observar en la Figura 4.21

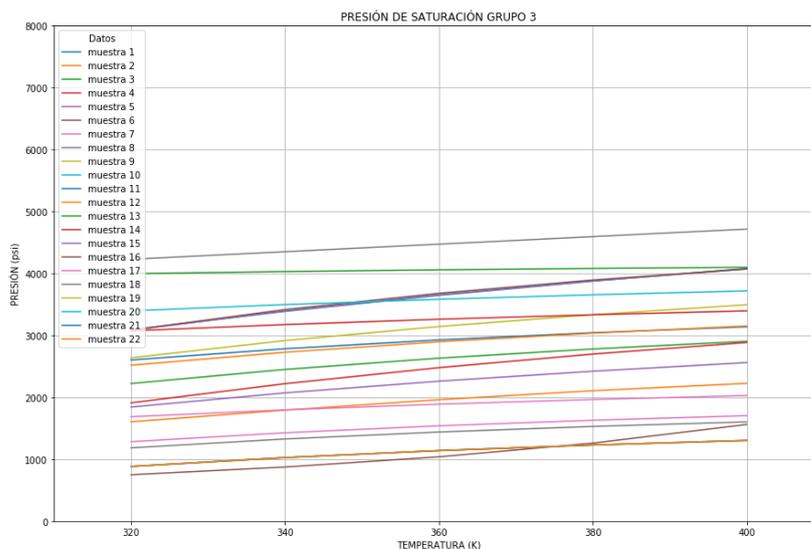


Figura 4.21 Tendencia de la presión de saturación del grupo 3 con respecto a la variación de temperatura.

En cuanto a la composición, en la Figura 4.22 se evidencia que la composición molar de los componentes no hidrocarburos es alto al igual que el del metano. Otro aspecto es que las fracciones pesadas de los crudos en su mayoría están por debajo del 40%, lo que indica que predominan los componentes livianos volátiles.

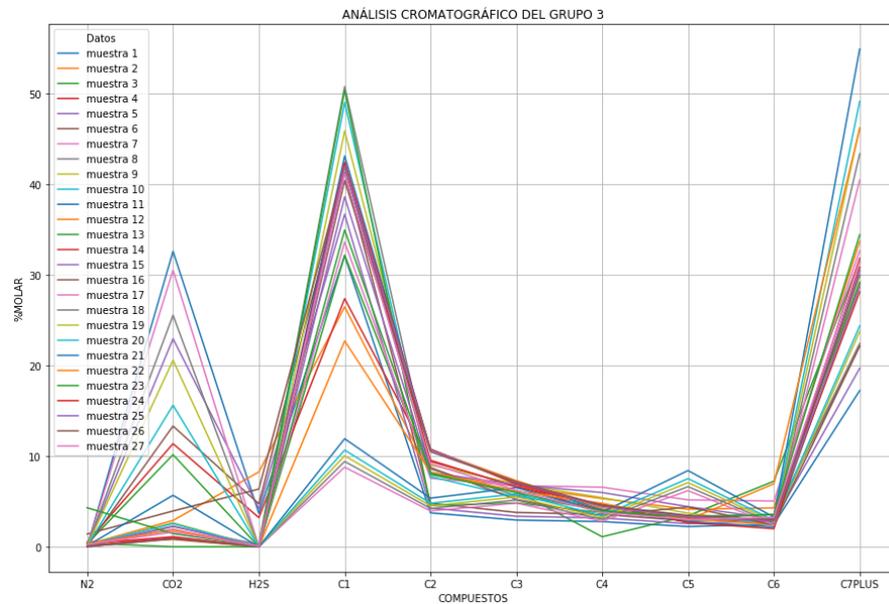


Figura 4.22 Análisis cromatográfico del grupo 3.

En la Figura 4.23 se muestra un alto contenido de saturados obtenido mediante el análisis SARA.

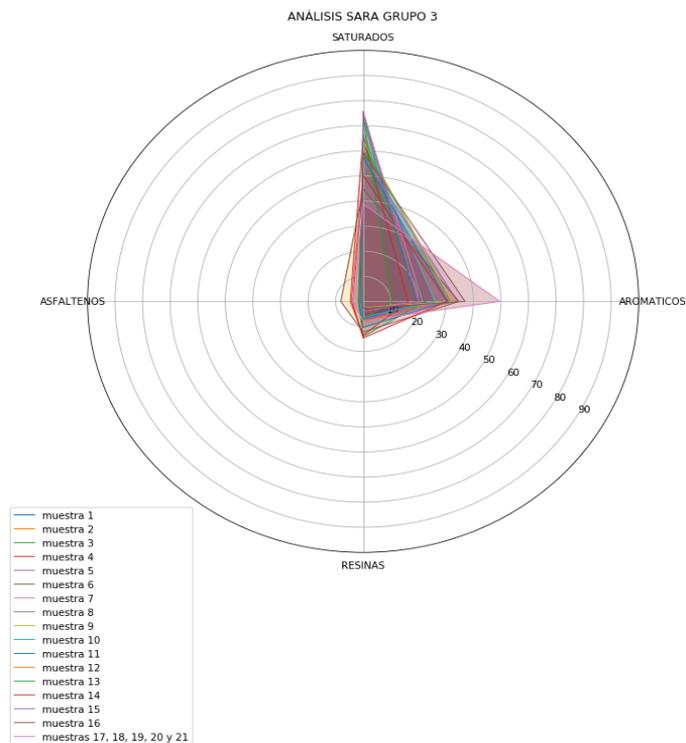


Figura 4.23 Análisis SARA del grupo 3.

4.3.4. Grupo 4

A este grupo pertenecen las muestras cuya presión Onset maneja rangos por encima de 7500 a cualquier temperatura, de igual manera con una pendiente negativa, este comportamiento se puede evidenciar en la Figura 4.24.

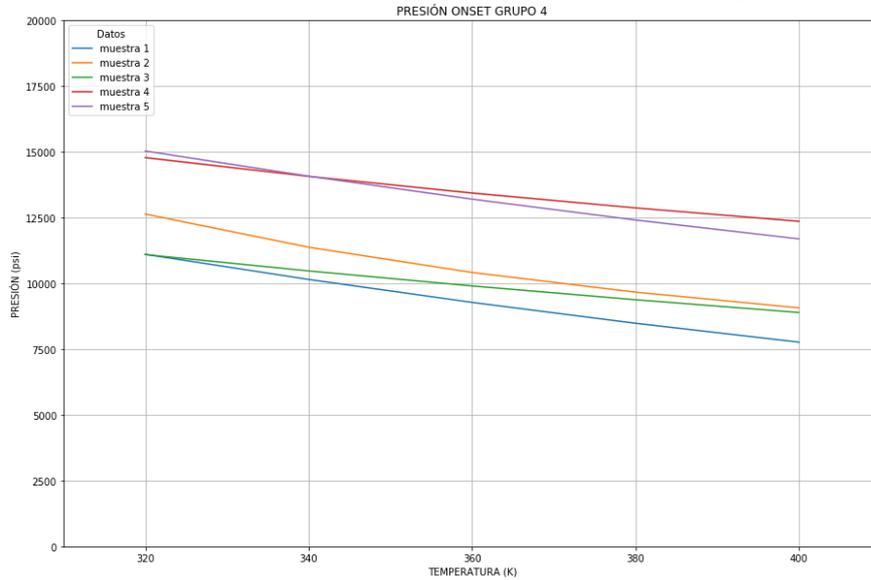


Figura 4.24 Gráfica AOP vs Temperatura Grupo 4.

En la Figura 4.25 se muestran las tendencias de las presiones de burbuja de las muestras incluidas en el último grupo.

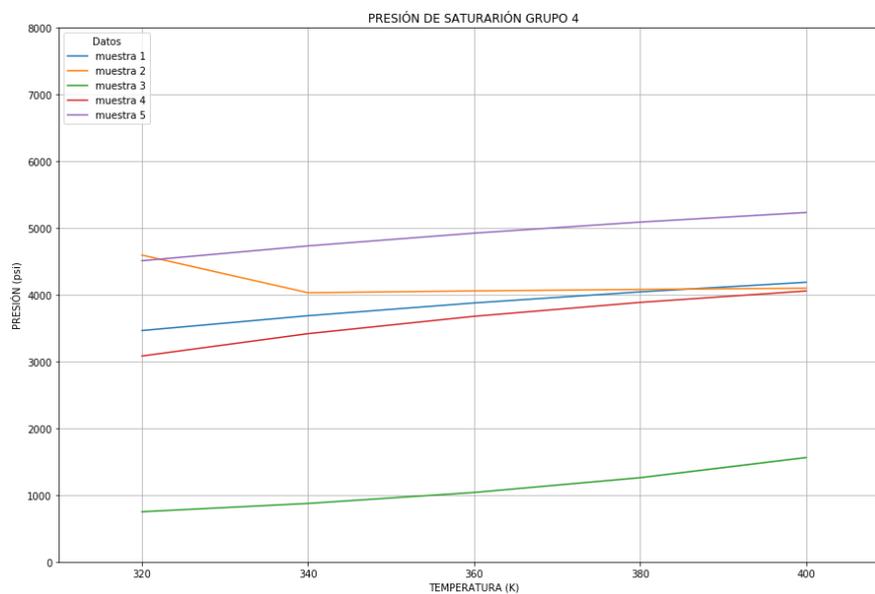


Figura 4.25 Tendencia de la presión de saturación del grupo 4 con respecto a la variación de temperatura.

En la Figura 4.26 se observa que, en cuanto a la composición molar, el metano está por encima del 40%, los componentes no hidrocarburos están por debajo del 4% y el C7+ está por debajo del 30%.

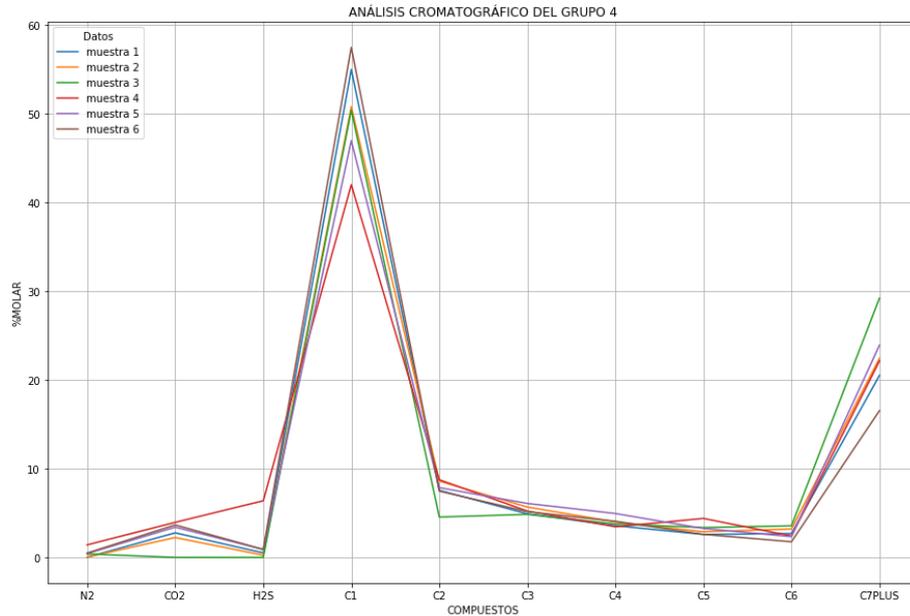


Figura 4.26 Análisis cromatográfico del grupo 4.

En la Figura 4.27, es claro que el % de asfaltenos determinado por el análisis SARA es muy bajo casi nulo, por tal motivo requiere de presiones muy altas para que una pequeña cantidad de este compuesto se precipite. De allí que los datos de presión onset obtenidos sean los más altos comparados con los otros grupos.

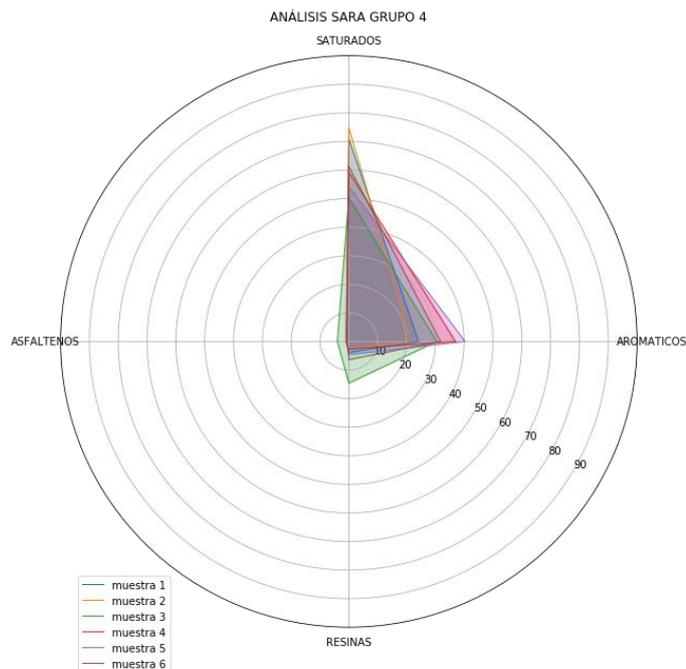


Figura 4.27 Análisis SARA del grupo 4.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES.

- Se logro desarrollar una red neuronal en el lenguaje de programación Python y con un modelo basado en estadísticos. La red neuronal tuvo un desempeño aceptable a la hora de predecir la presión onset en los yacimientos de petróleos mediante la composición química, el análisis SARA, la temperatura y la presión de saturación.
- El número de muestras en la semilla fue posible incrementarlo con la finalidad de obtener un mejor rendimiento a la hora de entrenar la red neuronal. Además, las nuevas muestras ingresadas a la base de datos presentan un comportamiento o una tendencia diferente a los crudos que inicialmente se tenían.
- La red neuronal tuvo un desempeño aceptable implementando una arquitectura o topología totalmente conectada; con esto es posible obtener todos los caminos existentes hacia la neurona de salida y así se pudo encontrar o calcular la relación no lineal entre el conjunto de entrada y el conjunto de salida, al mismo tiempo los estadísticos que se implementaron ayudaron a que la red neuronal tuviera una buena predicción aprendiendo de los errores que se obtenía en cada iteración de la etapa de entrenamiento.
- Se desarrollo un algoritmo de máquinas de soporte vectorial implementado en el mismo lenguaje de programación de la red neuronal AsphPy; con el objetivo de poder observar que método de predicción era más preciso a la hora de predecir la presión onset. Una vez calculado el rendimiento de la red neuronal y de las máquinas de soporte vectorial se pudo concluir que la red AsphPy con los mismos parámetros o conjunto de entrada tuvo mayor precisión teniendo un error promedio menor que al de las máquinas de soporte vectorial.
- El principal error de las muestras simuladas usando la red neuronal artificial se presentó en las muestras cuya presión onset supera los 8000 psi, las cuales en su mayoría pertenecen a las que contienen un alto contenido de gas.
- Las ecuaciones de estado preestablecidas para las simulaciones de las muestras en CMG, fueron inexactas cuando se estaban manejando temperaturas altas, por encima de 380 K. Inclusive la ecuación de Peng Robinson 1976, caracterizada por presentaba un error significativo cuando se intentó ajustar la presión de burbuja en el equilibrio liquido-vapor. Para una mejor calibración es recomendable incluir muchos más datos experimentales

- Los asfaltenos son moléculas muy complejas que requieren de mucho estudio en cuanto a su comportamiento a diferentes condiciones de presión y temperatura, actualmente no tienen una estructura definida ni peso molecular definido.
- Al momento de realizar el modelo de la red neuronal se debe escoger el número de iteraciones con mucha precaución, debido a que este parámetro define si el modelo de predicción diverge o converge. Cuando el número de iteraciones es bajo, es decir, la red se entrena con muy pocas iteraciones, el modelo será poco exacto a la hora de hacer predicciones. Si por el contrario a la red neuronal se le entrena con muchas iteraciones, su exactitud mejorará considerablemente; pero si se sobreentrena, es decir, el número de iteraciones es demasiado grande, la red neuronal tiende a perder convergencia y por lo tanto su exactitud bajará.

5.2. RECOMENDACIONES.

- Como recomendación, tener en cuenta cuál el tipo de agente precipitante a la hora de realizar el análisis SARA para la caracterización de los crudos, pues hay una diferencia significativa entre los asfaltenos de las muestras que son precipitados con diferentes n-alcános, para este caso, n-heptano y n-pentano. La relación existente H/C de los últimos es más alta, relacionándolo con su alto grado de aromaticidad.
- Con la finalidad de seguir mejorando el rendimiento de los algoritmos o modelos a la hora de predecir la presión onset en yacimientos de petróleos, se recomienda aumentar el número de muestras que se almacenan en el dataset para mejorar tanto la fase de entrenamiento o formulación de los modelos, como la validación de estos.
- Buscar otro tipo de arquitectura para la red neuronal y así observar cambios en el rendimiento a la hora de hacer las predicciones; al mismo tiempo cambiar parámetros para la formulación del modelo como la función de pérdidas, el optimizador y las métricas de evaluación.
- Cambiar y observar el rendimiento de las funciones que se implementan como kernel para la elaboración del modelo en las máquinas de soporte vectorial.
- Elaborar e implementar otros tipos de algoritmos de Machine Learning para la predicción de las presiones onset y así poder observar que algoritmo tiene mejor rendimiento o precisión.

- Determinar si existen otros parámetros o propiedades termodinámicas diferentes a las utilizadas en los trabajos anteriores en las muestras de crudo, que incidan o afecte la precipitación de los asfaltenos.
- Formular una interfaz gráfica amigable con el operador, para la implementación de los modelos de predicción en el sector de los hidrocarburos.
- Analizar y determinar que parámetros son más relevantes a la hora de calibrar las curvas por medio de regresiones líquido-vapor realizadas en el software CMG. Y de ser posible hacer uso de otros programas o plataformas con modelos más actualizados de calibración. Siempre teniendo en cuenta que la correlación que se aplica en cada caso sea la que se ajuste a la simulación de cada muestra.

Bibliografía

- Fernández Salgado & Yatzirih. (9 oct. 2017). *Caracterización de las fracciones SARA de crudos provenientes de hidrodesintegración*. Instituto Politécnico Nacional.
- Abedini A. Ashoori S. & Saki Y. (2010). *Application of Neural Network Model for Prediction of Asphaltene Precipitation*. Society of Petroleum Engineers.
- Ahmadi M. A. & Shadizadeh S. R. (2012). *New approach for prediction of asphaltene precipitation due to natural depletion by using evolutionary algorithm concept*.
- Ali Ahmadi M. & Golshadi M. (2012). *Neural network based swarm concept for prediction asphaltene precipitation due to natural depletion*. Journal of Petroleum Science.
- Alimohammadi S., Sayyad Amin J. & Nikooee E. (2017). *Estimation of asphaltene precipitation in light, medium and heavy oils: experimental study and neural network modeling*. Neural Computing and Applications.
- Al-Qasim, A., & Bubshait, A. (2017, April 23). *Asphaltenes: What Do We Know So Far*. Society of Petroleum Engineers. Society of Petroleum Engineers.
- Cano C. S. & Herrera S. (2016). *Desarrollo de un modelo para la determinación de la presión de inicio de precipitación de asfaltenos (Onset) en yacimientos de petróleos*.
- Cristiam David Cundar Paredes. (2013). *Modelamiento Escalado del Daño por Precipitación-Depositación de Asfaltenos en un Yacimiento de Petróleo*. Medellín, Colombia.
- E. F. Gholoum, G. P. Oskui & M. Salman. (2003). *Investigation of Asphaltene Precipitation onset Conditions for Kuwaiti Reservoirs*.
- Matich, D. J. (Marzo de 2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Rosario.
- Mohammadi A. H. & Richon D. (2008). *Estimating Onset of Precipitation of Dissolved Asphaltene in the Solution of Solvent + Precipitant Using Artificial*. Neural Network Technique.
- Olave N. F. & Bahamon N. H. Q. (2018). *DETERMINACIÓN DE LA PRESIÓN ONSET USANDO ANÁLISIS ESTADÍSTICO ANOVA EN MODELO DE RED NEURONAL*. Neiva.
- Padilla D. & Watt K. (2012). *Precipitación de asfaltenos: Técnicas de predicción y control Asphaltene precipitation: Prediction and Control Techniques*.
- PÉREZ, C. D. (2015). *Modelo de estabilidad de asfaltenos como herramienta para predecir el daño de formación en pozos productores de petróleo con alto contenido de CH₄, CO₂ o N₂*. Medellín, Colombia: Universidad Nacional de Colombia.
- Súarez, E. J. (2013). *Tutorial sobre Máquinas de Vectores Soporte (SVM)*. Madrid, España.
- Zarei F. & Baghban A. (2017). *Phase behavior modelling of asphaltene precipitation utilizing MLP-ANN approach*. Petroleum Science and Technology.

Zendehboudi S., Ahmadi M. A., Mohammadzadeh O., Bahadori A. & Chatzis I. (2013). *Thermodynamic investigation of asphaltene precipitation during primary oil production: Laboratory and smart technique*. Industrial and Engineering Chemistry Research.

«Sinapsis: ¿sabes cómo se conectan tus neuronas?» 2014. *neuronas en crecimiento*. <https://neuropediatra.org/2014/06/04/sinapsis-neuronal/> (27 de agosto de 2019).

Ltd., C. M. (2015). *WINPRO USER GUIDE, Phase-Behaviour & fluid property program*. Calgary, Canada.

Peyman Bahrami, R. K. (2015). Prediction of the gas injection effect on the asphaltene phase envelope. *OGST*, 13.

ANEXOS

COMPARACIÓN ENTRE LA PRESIÓN ONSET REAL DE 100 MUESTRAS ALEATORIAS, LOS PREDICHOS CON ASPHPY Y EL SVM.

Presión onset real (psi)	Presión Onset de las máquinas de soporte vectorial. (psi)	Presión onset predicción AsphPy (psi)	Error REAL VS SVM (%)	Error REAL VS RNA (%)
288.6384277	303.1303	317.5684814	5.020766	10.02294
444.3517151	5931.401	461.2319031	1234.844	3.798835
664.5075684	678.9971	690.6247559	2.180495	3.930307
705.2028198	719.6699	721.2747803	2.051476	2.279055
762.7471313	777.2517	776.3048706	1.901623	1.777488
819.3587036	833.87	819.1177979	1.771064	0.029402
834.4080811	848.9114	863.0143433	1.738165	3.42833
925.649231	5816.384	1186.377441	528.3573	28.16706
1045.085693	1059.579	1065.718872	1.386779	1.974305
1119.903687	1134.469	1107.975952	1.300593	1.065068
1177.447998	5644.858	1071.701416	379.4147	8.980998
1249.108887	5825.472	1046.916626	366.3703	16.18692
1285.043213	1299.535	1270.549805	1.127731	1.127854
1340.350098	1354.908	1343.805908	1.086106	0.257829
1369.032471	5765.789	987.7965698	321.1579	27.8471
1459.786499	1474.286	1459.070313	0.993274	0.049061
1492.639893	1507.134	1497.304199	0.971058	0.312487
1866.624512	1881.127	1855.589478	0.776962	0.591176
2023.748779	2038.207	2010.98291	0.714442	0.630803
2043.373291	4604.398	1884.438965	125.3332	7.778037
2100.917725	2115.41	2109.238525	0.689815	0.396056
2172.578613	4705.255	2151.430908	116.5747	0.973392
2203.734619	2218.292	2149.057617	0.660596	2.481106
2232.655029	5931.401	2823.351318	165.6658	26.45712
2263.81958	2278.286	2310.828369	0.63903	2.076525
2351.395508	2365.94	2374.324707	0.618533	0.975132
2383.256104	2397.808	2433.783203	0.610588	2.120087
2401.701416	2416.198	2477.890137	0.603607	3.172281
2412.955322	2427.507	2425.135498	0.603068	0.504782
2414.324707	2428.821	1601.334839	0.600435	33.67359
2461.298584	2475.8	2460.108398	0.58917	0.048356
2504.521973	2519.032	2532.712891	0.579348	1.125601
2528.260742	2542.82	2277.754639	0.575884	9.908238
2528.260742	2542.756	2628.932373	0.573344	3.981853
2578.518311	3549.536	2632.802734	37.658	2.105256
2614.585938	2629.025	2702.12793	0.552256	3.348216
2617.741211	2632.24	2642.569336	0.553858	0.948456
2632.943848	2647.469	2673.774902	0.551664	1.550776

2657.397705	2671.895	2704.33667	0.545567	1.766351
2665.941162	2680.508	2735.53418	0.546416	2.610448
2765.941162	5830.454	2466.1875	110.7946	10.83731
2791.450928	2805.94	2386.25415	0.519057	14.51563
2791.450928	5728.832	2715.336182	105.2278	2.726709
2803.849121	2818.407	2831.700439	0.519195	0.993324
2831.50415	2846.062	2926.181396	0.514126	3.343709
2909.419434	2923.915	2907.723633	0.498225	0.058287
2931.504395	5835.575	2767.159668	99.06419	5.606157
2967.02832	2981.586	3012.087646	0.490652	1.518669
2971.438965	2985.93	3070.485107	0.487699	3.333272
2992.524902	3640.143	3085.721436	21.64118	3.114311
3026.458984	3040.96	3028.793945	0.479145	0.077152
3028.261719	5824.108	3189.537109	92.32515	5.325675
3056.834961	3071.363	3147.454346	0.475261	2.964484
3126.708984	3141.21	3075.358643	0.463795	1.642313
3148.303955	3162.795	2935.526855	0.460294	6.758468
3148.303955	3162.755	3133.33252	0.459021	0.47554
3156.835205	5842.599	3145.791016	85.07774	0.34985
3201.106201	3215.601	3330.027832	0.45282	4.027409
3209.246094	3223.691	3280.992188	0.450118	2.235606
3237.69873	3252.162	3268.996338	0.446707	0.966662
3291.451904	3305.903	3276.665771	0.43904	0.449228
3379.694824	3394.202	3480.125244	0.42925	2.971582
3473.022217	3487.508	3434.9729	0.417104	1.095568
3479.694824	3494.259	3518.433594	0.418567	1.113281
3483.6875	3498.188	3610.516113	0.416256	3.640643
3489.584229	3504.073	3601.386475	0.415218	3.203884
3511.906006	3526.403	3594.005615	0.412799	2.337751
3541.709717	3556.166	4170.273438	0.408178	17.74747
3581.365234	3800.811	3474.778076	6.127428	2.97616
3604.412842	3618.919	3753.206543	0.402465	4.128098
3610.521973	5736.791	3792.379639	58.89092	5.03688
3611.875	5645.021	4387.952148	56.29058	21.48682
3617.549805	3632.061	3750.925781	0.401132	3.686915
3624.780518	4103.832	3632.550537	13.21601	0.214358
3630.436279	3644.895	3590.960449	0.398262	1.087358
3648.304932	3662.863	3601.306152	0.399032	1.288236
3660.55249	5824.747	3675.360107	59.12206	0.404519
3660.55249	3675.107	3713.872803	0.397611	1.456619
3719.080322	3733.583	3820.471191	0.389961	2.726235
3744.192139	3758.713	3769.06665	0.387812	0.664349
3762.405762	5931.401	3379.435791	57.64916	10.17886
3784.586182	3799.096	3791.472656	0.383397	0.181961
3801.853516	3816.35	3812.418945	0.381313	0.277902
3813.79248	5692.868	1504.732544	49.27054	60.54498
3819.802002	3834.357	4675.076172	0.381028	22.39054
3877.766357	3892.271	3867.224365	0.374042	0.271857
3903.537598	3918.095	3929.765137	0.372929	0.671892
3906.173828	3920.67	3976.321045	0.371119	1.795804
3952.119141	3966.642	3908.879395	0.367474	1.09409

3977.766602	3992.325	4009.953369	0.365997	0.809167
4003.399658	4017.828	4136.852051	0.360406	3.333477
4072.507568	5754.625	4897.437012	41.30422	20.25606
4089.239502	5931.401	3338.806152	45.04901	18.35142
4094.967285	4109.421	4053.587646	0.352963	1.0105
4127.349609	4141.833	4307.621582	0.350904	4.367742
4129.803711	5931.401	4320.73877	43.62429	4.623345
4145.616211	5931.401	4354.79248	43.07646	5.045722
4160.553711	4175.042	4150.241699	0.34824	0.247852
4204.824707	4219.253	4235.790527	0.343158	0.736435
4262.279297	5543.928	4856.559082	30.06957	13.94277

COMPOSICIÓN QUÍMICA DE LAS MUESTRAS SIN NORMALIZAR.

N2	CO2	H2S	C1	C2	C3	C4	C5	C6	C7Plus
0.06	32.58	3.69	32.09	3.75	2.95	2.78	2.23	2.49	17.22
0.27	2.9	8.26	26.45	8.04	6.66	5.32	4.14	4.29	33.66
0.33	10.16	0	32.16	7.97	6.14	1.1	3.42	7.23	31.49
0.49	11.37	3.22	27.36	9.41	6.7	3.98	3.2	2.49	31.79
0.07	22.94	4.22	36.67	4.28	3.37	3.18	2.55	2.85	19.68
0.34	0.16	0	32.9	8.15	6.28	4.63	3.7	4	39.85
0.37	0.17	0	35.73	8.85	6.82	1.23	3.8	8.04	34.98
0.08	13.31	4.74	41.26	4.82	3.79	3.57	2.87	3.2	22.14
0.16	1.94	0	33.6	7.56	6.74	6.58	5.18	5.05	32.63
0.02	2.76	0.51	54.92	7.55	4.9	3.57	2.58	2.68	20.49
0.04	2.23	0.25	50.74	8.59	5.66	4	2.88	3.2	22.42
0.24	2.64	0	45.86	9.01	6.93	5.42	3.71	2.48	23.73
0.24	2.56	0	48.99	7.74	5.5	4.62	3.38	2.64	24.25
0.01	0.99	0	43.09	10.72	7.08	4.1	2.77	2.18	29.06
0.56	0.81	0	41.87	10.7	7.3	4.26	3.09	2.46	28.96
0.39	0.01	0	50.34	4.54	4.85	3.77	3.36	3.56	29.18
0.33	1.11	0	42.19	9.54	6.61	4.66	2.67	1.99	30.84
0.2	2.38	0	40.4	8.39	6.87	5.85	4.25	2.92	28.77
0.56	0.13	0	46.9	6.13	4.67	3.05	2.25	2.5	33.82
0.19	2.29	0	38.59	8.18	6.85	5.99	4.43	3.07	30.44
1.28	1.15	0	35.98	10.3	8.44	4.83	2.33	2.36	33.34
1.41	3.95	6.37	41.95	8.74	5.21	3.45	4.39	2.41	22.12
0.18	2.2	0	36.77	7.98	6.83	6.13	4.61	3.35	32
0.17	2.11	0	34.95	7.77	6.81	6.27	4.79	3.5	33.67
0.16	1.94	0	33.53	7.66	7.27	6.58	5.18	5.05	32.66
0.08	30.5	0	8.78	3.95	4.81	2.82	6.2	2.42	40.44
0.09	25.54	0	9.41	4.24	5.15	3.02	6.64	2.59	43.33
0.09	20.58	0	10.03	4.52	5.5	3.22	7.08	2.76	46.22
0.1	15.61	0	10.66	4.8	5.84	3.43	7.52	2.93	49.1

0.1	10.65	0	11.29	5.08	6.18	3.63	7.97	3.11	51.99
0.11	5.68	0	11.92	5.37	6.53	3.83	8.41	3.28	54.88
0.12	0.72	0	12.54	5.65	6.87	4.03	8.85	3.45	57.77
0.3	1.83	0	22.7	8.24	6.14	4.8	2.97	6.95	46.09
5.08	11.12	0.06	30.99	6.41	5.04	3.56	3	3.26	31.48
0.19	9.34	0.06	31.39	7.43	5.73	3.94	3.32	3.61	34.99
0.19	9.76	0.06	32.34	7.57	5.71	3.85	3.21	3.49	33.82
0.24	1.67	0.06	27.9	8.99	8.65	5.39	3.96	4.07	39.07
0.2	1.43	0.06	33.01	7.91	6.23	4.37	3.7	4.03	39.07
4.31	1.45	0.07	34.94	8.09	5.98	3.95	3.27	3.55	34.4
0.09	1.02	0.05	42.41	10.78	6.92	4.47	3.29	2.86	28.11
0.03	0.89	0.03	41.95	10.68	7.11	4.44	3.32	2.89	28.66
0.04	0.78	0	40.91	10.72	6.64	4.44	2.62	2.51	31.34
0.04	0.86	0.03	40.38	10.48	7.17	4.58	3.47	3.02	29.95
0.24	1.62	0.09	41.13	9.16	6.23	3.78	3.02	3.25	31.49
0.3	1.97	0.08	34.01	10.67	9.59	5.21	3.38	3.31	31.49
0.23	1.58	1.74	39.89	8.92	6.12	3.76	3.01	3.25	31.49
0.1	3.71	1.85	46.53	8.76	4.98	3.32	1.92	2.64	26.19
0.48	0.92	0	43.43	11.02	6.55	4.49	3.53	2.7	26.88
0.8	0.05	0	51.02	8.09	6.02	3.97	3.21	2.67	24.17
0.21	2.56	0	25.2	7.08	6.56	6.2	4.92	3.94	43.33
0.44	3.4	0.9	46.91	7.86	6.07	4.94	3.26	2.33	23.89
0.37	3.23	0.3	48.66	6.73	4.97	4.44	3.34	2.51	25.45
0.49	3.65	0.9	57.41	7.46	5.15	4.07	2.59	1.76	16.52
0.26	17.57	0	37.65	5.41	4.15	3.93	3.15	2.46	25.43
0.15	0.27	0	26.79	6.8	8.37	5.97	3.95	5.47	42.23
0.06	2.45	0.59	38.65	6.66	5.33	3.93	2.75	4.67	34.92
0.21	5.14	2.7	22	7.1	5.34	3.77	2.53	5.55	45.66
0.13	2.07	0.02	17.06	8.05	6.88	5.76	5.17	3.35	51.52
0.390	1.740	0.000	20.550	7.310	5.340	4.650	7.850	5.480	46.690
0.371	1.653	0.000	24.523	6.945	5.073	4.418	7.458	5.206	44.356
0.351	1.566	0.000	28.495	6.579	4.806	4.185	7.065	4.932	42.021
2.880	1.697	0.000	20.036	7.127	5.207	4.534	7.654	5.343	45.523
5.371	1.653	0.000	19.523	6.945	5.073	4.418	7.458	5.206	44.356
7.861	1.610	0.000	19.009	6.762	4.940	4.301	7.261	5.069	43.188
0.351	11.566	0.000	18.495	6.579	4.806	4.185	7.065	4.932	42.021
0.293	26.305	0.000	15.413	5.483	4.005	3.488	5.888	4.110	35.018
0.195	50.870	0.000	10.275	3.655	2.670	2.325	3.925	2.740	23.345
0.416	1.858	0.000	21.942	7.806	5.695	4.843	7.685	5.328	44.429
0.442	1.976	0.000	23.333	8.301	6.049	5.036	7.520	5.176	42.167

PRESION ONSET Y PRESION DE SATURACIÓN A 5 TEMPERATURAS.

Pb (psi)	Ponset(psi)
3089.22248	9002.28462
3387.04656	7932.01643
3645.62682	7173.88931
3874.37199	6608.72896
4079.54267	6171.17899
1603.66608	5073.78585
1791.99563	4492.50623
1960.4676	4129.80378
2105.47545	3903.53771
2224.33621	3762.40584
2221.3934	8081.34894
2449.18871	7049.34852
2630.93239	6436.15938
2779.52811	6043.06428
2903.45292	5779.28871
1909.73052	8578.67274
2218.83055	6767.26022
2477.85938	5400.3179
2697.07547	4353.63922
2884.23862	3541.70977
3080.06356	8037.80084
3396.25891	7124.2411
3657.48269	6439.14262
3878.75136	5911.0759
4069.59352	5494.63239
1951.18962	5396.69474
2138.82752	4787.23898
2274.51247	4661.90375
2370.1714	4636.12842
2436.67811	4630.8277
1924.65084	5916.68427
2129.50357	5141.87348
2312.42076	4754.54725
2478.5303	4520.21
2631.2379	4362.62215
3084.53574	7475.4349
3416.49025	6127.33239
3678.98994	5134.55168
3893.03726	4383.43486
4071.42381	3801.85354

Pb(psi)	Ponset(psi)
1685.47793	4290.16583
1796.91957	3473.02232
1887.82066	2909.41954
1963.50005	2504.52198
2027.51777	2203.73473
3463.40062	11105.7111
3685.31886	10149.1938
3876.59526	9276.51043
4042.25304	8483.63826
4186.43171	7765.39366
4227.51772	14048.6335
4349.73582	11589.4925
4471.61965	9845.84212
4593.3695	8564.50525
4715.1638	7594.54129
2636.73729	11254.3233
2915.1862	9657.0281
3142.04837	8465.17178
3331.96296	7541.38968
3494.07148	6804.10697
3394.57851	10557.3232
3496.27412	9185.07174
3581.72466	8045.46357
3654.45694	7089.77324
3717.04296	6281.29892
2601.6719	10482.5949
2782.72524	8453.71134
2925.83398	7048.64781
3041.6911	6146.98755
3137.22393	5655.53956
2517.02352	12058.3917
2726.5195	9668.75723
2895.16856	7896.19015
3034.30656	6680.46239
3151.22461	5969.96681
3996.58847	12636.0642
4029.66094	11374.3839
4056.89892	10414.6407
4079.33171	9665.95773
4097.80701	9069.48881

Pb(psi)	Ponset(psi)
3071.9991	7076.5722
3173.69471	6424.66091
3259.14525	5932.87187
3331.87753	5548.50442
3394.46355	5239.71257
1938.44157	4920.37351
2160.34219	4518.90566
2348.36649	4204.82462
2509.14508	3952.11915
2647.75249	3744.19222
3292.98484	5635.1288
3451.22997	5249.20837
3561.25111	5064.38591
3629.15543	4955.82489
3662.18837	4884.31051
1841.73296	4940.42861
2070.14518	4094.96728
2259.08984	3581.36519
2419.92018	3237.69876
2559.56454	2992.5249
2069.29056	3489.58412
2257.04542	2967.02827
2414.38581	2632.94394
2547.54298	2401.70142
2661.23041	2232.65502
748.873051	11094.6498
873.221443	10472.788
1037.93438	9902.33504
1259.02594	9377.34149
1560.30211	8892.76167
1744.35344	3624.78046
1902.83489	3126.70887
2039.61677	2803.84921
2158.64224	2578.51825
2262.97879	2412.95522
1553.08925	2383.25613
1680.20978	2263.81968
1796.58402	2172.57853
1903.64544	2100.91763
2002.57335	2043.37332

1667.29153	4282.75763
1811.029	3483.68753
1923.22686	2971.4389
2013.30091	2614.58584
2087.13064	2351.39558
1282.55305	4959.62374
1426.29052	4160.55363
1538.48839	3648.30501
1628.56243	3291.45194
1702.39217	3028.26168
1182.55285	4459.6227
1326.29032	3660.5526
1438.48818	3148.30397
1528.56222	2791.4509
1602.39196	2528.26065
882.552225	4459.6227
1026.2897	3660.5526
1138.48756	3148.30397
1228.5616	2791.4509
1302.39134	2528.26065
384.185141	4958.62453
684.173258	3610.52202
951.04906	2617.74132
1172.5764	1866.6245
1359.46282	1285.04317
574.395369	1459.78652
732.876819	1340.35006
869.658693	1249.10892
988.684168	1177.44802
1093.02071	1119.90371
213.669802	3813.79251
454.581513	2414.3246
700.068415	1369.03252
921.691537	664.50759
1112.281	288.638433
313.394828	1045.08566
471.876278	925.649206
608.658152	834.408058
727.683628	762.747159
832.020173	705.202848
696.494629	5630.95895
1031.13778	4918.69047
1290.92941	4285.56294

1499.32081	3719.08041
1670.39015	3209.24613
3381.25443	6288.05949
3603.15505	5886.59165
3791.17934	5572.5106
3951.95793	5319.80513
4090.56534	5111.8782
2181.25194	4788.05638
2403.15256	4386.58854
2591.17686	4072.50749
2751.95545	3819.80203
2890.56286	3611.8751
2681.25298	5888.05866
2903.1536	5486.59082
3091.17789	5172.50977
3251.95648	4919.8043
3390.56389	4711.87738
1924.39816	3977.76651
2082.87961	3479.69491
2219.66149	3156.83526
2338.68696	2931.5043
2443.02351	2765.94126
1824.39796	3877.7663
1982.87941	3379.69471
2119.66128	3056.83505
2238.68676	2831.50409
2343.0233	2665.94106
3624.40168	5877.77044
3782.88313	5379.69885
3919.66501	5056.83919
4038.69048	4831.50823
4143.02703	4665.9452
2675.63651	11783.0109
2885.1325	9393.3764
3053.78155	7620.80933
3192.91956	6405.08157
3309.8376	5694.58599
2891.87875	9493.38174
2993.57436	8121.13023
3079.0249	6981.52206
3151.75718	6025.83174
3214.3432	5217.35741
2794.28809	6930.60553

2952.53322	6544.6851
3062.55436	6359.86264
3130.45868	6251.30162
3163.49161	6179.78724
2766.8785	9392.38153
2868.57411	8020.13002
2954.02464	6880.52185
3026.75692	5924.83153
3089.34294	5116.3572
3344.63433	6417.74772
3532.38918	5895.19187
3689.72957	5561.10754
3822.88674	5329.86502
3936.57417	5160.81862
2758.5418	8177.75706
2862.67334	6604.87223
2949.93659	5477.80127
3023.78747	4641.68792
3086.84016	4003.39965
3464.63458	6317.74751
3652.38943	5795.19166
3809.72982	5461.10733
3942.88699	5229.86481
4056.57442	5060.81841
3041.08951	9142.90406
3263.00775	8186.38676
3454.28416	7313.70335
3619.94193	6520.83118
3764.12061	5802.58658
3001.64481	10062.4822
3105.77635	8489.59733
3193.0396	7362.52637
3266.89048	6526.41302
3329.94317	5888.12474
4717.37172	8548.86894
5026.47175	6737.45643
5285.50058	5370.5141
5504.71667	4323.83543
5691.87981	3511.90598
3675.86677	4527.07758
3984.96679	2657.39766
4243.99563	1492.63988
4463.21172	819.358676

4650.37486	444.351702	1106.43194	8336.25783	2191.3002	5957.4802
3081.98369	14780.7363	1366.22358	7703.13029	2236.6977	4686.3294
3416.62683	14068.4678	1574.61498	7136.64776	2449.6050	7769.8466
3676.41847	13435.3403	1745.68432	6626.81348	2497.8464	8674.4951
3884.80987	12868.8578	410.432826	11398.9985	2546.5426	8231.8705
4055.87921	12359.0235	688.677961	9587.58595	2575.8907	6967.1479
3586.47861	8797.33869	943.08066	8220.64363	2605.9334	5473.8673
3761.29653	8142.18418	1170.20026	7173.96495	1272.2408	3906.1737
3905.89236	7663.89451	1366.6905	6362.0355	1419.3490	4835.6946
4027.49545	7300.93922	1173.0547	4127.3498	1577.5246	4945.6792
4131.14844	7017.20251	1297.9214	5033.2411	1701.9186	4454.3675
4509.09255	15031.9163	1402.8071	4958.0922	1806.6243	3722.4979
4731.01079	14075.3989	1493.2915	4262.2793	1395.5212	3604.4127
4922.28719	13202.7155	1574.2569	3323.8246	1616.4971	4656.4650
5087.94497	12409.8434	1429.0042	5505.2112	1826.3212	5004.6377
5232.12365	11691.5988	1567.2779	6377.3220	2018.9249	4808.3701
3124.41453	4854.85436	1684.2785	6187.7794	2178.3166	4315.3361
3422.23861	3784.58617	1782.7308	5320.2012	1691.9970	3630.4362
3680.81886	3026.45905	1862.9556	4205.9702	2087.6969	4796.5955
3909.56403	2461.2987	1693.8361	7112.3244	2446.3460	5513.1848
4114.73472	2023.74873	1856.6224	7968.8801	2780.8309	5797.0829
2263.26699	9126.20257	1980.3920	7653.1495	3065.8377	5752.5830
2572.36702	7314.79005	2072.5521	6587.0312	1287.6840	5217.5097
2831.39585	5947.84773	2170.0953	5242.9151	1428.7371	6127.0474
3050.61194	4901.16906	1566.1618	5165.0324	1536.5193	5960.4960
3237.77508	4089.2396	1672.9506	6090.3338	1643.1330	5137.9600
2191.28632	5744.27466	1716.4661	5920.4900	1720.0774	4051.8486
2507.48167	4830.71491	1855.1331	5067.6088	1416.4503	6598.0668
2768.70546	4145.61643	1904.9778	3970.6531	1568.1026	7480.8647
2989.97412	3617.54971	2003.7064	6381.0227	1683.5047	7186.4019
3180.81628	3201.10621	2084.6962	7295.9948	1799.7166	6196.5368
771.788797	9048.5263	2140.9341	6991.5197	1889.7230	4920.9109

RED NEURONAL Y SUPPORT VECTOR MACHINE CODIGO.

```

from keras.models import Sequential
from keras.layers.core import Dense
from keras import metrics
from sklearn.svm import SVR
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
import pandas as pd

```

```

from pandas import ExcelWriter
import numpy as np
np.random.seed(3)
predictoras = pd.read_excel("../datasets/predictoras.xlsx")
ponset = pd.read_excel("../datasets/ponset.xlsx")
Muestras = pd.concat([predictoras,ponset], axis = 1)
Muestras.head()
scaler=MinMaxScaler()
X_vals = np.array(predictoras, "float32")
Y_vals = np.array(Muestras["Ponset"],"float32")
print(X_vals.shape,Y_vals.shape, Muestras.shape)
X_vals = scaler.fit_transform(X_vals)
train_id = np.random.choice(len(X_vals),round(len(X_vals)*0.8), replace=False)
test_id= np.array(list(set(range(len(X_vals)))-set(train_id)))
len(train_id)
x_vals_train = X_vals[train_id]
y_vals_train = Y_vals[train_id]
x_vals_test = X_vals[test_id]
y_vals_test = Y_vals[test_id]
model = Sequential()
model.add(Dense(50, input_dim=16, activation='relu'))
model.add(Dense(45, input_dim=50, activation='relu'))
model.add(Dense(40, input_dim=45, activation='relu'))
model.add(Dense(35, input_dim=40, activation='relu'))
model.add(Dense(30, input_dim=35, activation='relu'))
model.add(Dense(25, input_dim=30, activation='relu'))
model.add(Dense(20, input_dim=25, activation='relu'))
model.add(Dense(15, input_dim=20, activation='relu'))
model.add(Dense(1, activation='relu'))
model.compile(loss='mean_absolute_percentage_error',
              optimizer='adam',
              metrics=['mae'])
history = model.fit(x_vals_train, np.transpose([y_vals_train]), epochs=100000,
batch_size=50,validation_split=0.1)
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper right')
plt.show()
scores = model.evaluate(x_vals_train, np.transpose([y_vals_train]))
train = pd.DataFrame(x_vals_train)
test = pd.DataFrame(x_vals_test)
datos = pd.concat([train,test], axis = 0)
y_train = pd.DataFrame(y_vals_train)
y_test = pd.DataFrame(y_vals_test)

```

```

onset = pd.concat([y_train,y_test],axis = 0)
Onset_real = np.array(onset)
Datos = np.array(datos)
y_pre = model.predict([Datos])
prediccion = pd.DataFrame(y_pre, columns=["Ponset_pre"])
onset_tes = pd.DataFrame(Onset_real, columns = ["Ponse_real"])
datos=pd.DataFrame(Datos,columns=["Temperatura","N2","CO2","H2S","C1","C2","C3","C4","C5","C6","C7Plus","SATURADOS","AROMATICOS","RESINAS","ASFALTENOS","Pb"])
Data_set = pd.concat([datos, onset_tes,prediccion], axis = 1)
Data_set.head()
eje_x = list(range(334))
plt.figure(figsize=(15,5))
plt.plot(eje_x,sorted(Y_vals),'r-',label='Ponset Real')
plt.plot(eje_x,sorted(y_pre),'g-',label='Ponset RNA')
plt.legend(title='Datos',loc="upper left")
plt.title("Comparación: Presión Onset REAL y AsphPy")
plt.ylabel("Presión Onset (psi)")
plt.xlabel("Numero de Muestra")
plt.grid(True)
plt.show
Error_RNA = np.mean((np.abs((Onset_real-y_pre)/Onset_real)*100))
Error_RNA
writer = ExcelWriter("../datasets/DATOS_RED_NEURONAL.xlsx")
Data_set.to_excel(writer, 'Ponset', index=False)
writer.save()
# SUPPORT VECTOR MACHINE
predictoras1 = pd.read_excel("../datasets/predictorasint.xlsx")
ponset1 = pd.read_excel("../datasets/ponsetsint.xlsx")
Muestras1 = pd.concat([predictoras1,ponset1], axis = 1)
Muestras1.head()
X_vals1 = np.array(predictoras1, "float32")
Y_vals1 = np.array(Muestras1["Ponset"],"float32")
print(X_vals1.shape,Y_vals1.shape, Muestras1.shape)
x_vals_train1 = X_vals1[train_id]
y_vals_train1 = Y_vals1[train_id]
C = 1e3
svr_rbf = SVR(kernel="rbf", C=C, gamma=0.1)
model_rbf = svr_rbf.fit(x_vals_train1,y_vals_train1)
Y_pre_rbf = model_rbf.predict(X_vals1)
Prediccion_rbf = pd.DataFrame((Y_pre_rbf*145.038), columns=["Ponset_rbf"])
onset_tes1 = pd.DataFrame((Y_vals1*145.038), columns = ["Ponse_real"])
datos1=pd.DataFrame(X_vals1,columns=["Temperatura","N2","CO2","H2S","C1","C2","C3","C4","C5","C6","C7Plus","SATURADOS","AROMATICOS","RESINA S","ASFALTENOS","Pb"])
data_set = pd.concat([datos1, onset_tes1,Prediccion_rbf], axis = 1)
data_set.head()

```

```

plt.figure(figsize=(15,5))
plt.plot(eje_x,sorted(Y_vals1*145.038),'r-',label='Ponset Real')
plt.plot(eje_x,sorted(Y_pre_rbf*145.038),'b-',label='Ponset SVM')
plt.legend(title='Datos',loc="upper left")
plt.title("Comparación: Presión Onset REAL y SVM")
plt.ylabel("Presión Onset (psi)")
plt.xlabel("Numero de Muestra")
plt.grid(True)
plt.show()
writer = ExcelWriter("../datasets/DATOS_SVM.xlsx")
data_set.to_excel(writer, 'Ponset', index=False)
writer.save()
Error_SVM=np.mean((np.abs(((Y_vals1*145.038)-
(Y_pre_rbf*145.038))/Y_vals)*100))
Error_SVM
# GRAFICO DE COMPARACION
plt.figure(figsize=(15,5))
plt.plot(eje_x,sorted(Y_vals),'r-',label='Ponset Real')
plt.plot(eje_x,sorted(Y_pre_rbf*145.038),'b-',label='Ponset SVM')
plt.plot(eje_x,sorted(y_pre),'g-',label='Ponset RNA')
plt.legend(title='Datos',loc="upper left")
plt.title("Comparación: Presión Onset REAL, AsphPy y SVM")
plt.ylabel("Presión Onset (psi)")
plt.xlabel("Numero de Muestra")
plt.grid(True)
plt.show()

```