Global Journal of Engineering and Technology [GJET].
ISSN: **2583-3359** (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

Three-Stage Machine Learning Pipeline Ensemble with Gradient Boosting and SHAP Analysis - Evaluating Flow Properties of Heavy Crude Under Thermal Conditions

BY

[1,2]Falade, A.A, [3,4] Olanrewaju Sa'id, & [5]Akinsete, O.O

[1]Department of Mining and Petroleum Resources Engineering, Federal Polytechnic Ado Ekiti,Nigeria
[2]Ph.D candidate ,department of Petroleum Engineering, University of Ibadan, Nigeria
[3]Department of Finance and Data Analytics, University of Stirling , Scotland
[4]Circle Data and IT Solutions Ltd, United Kingdom
[5]Department of Petroleum Engineering, University of Ibadan, Nigeria

Article History

Received: 15/01/2025
Accepted: 28/01/2025
Published: 31/01/2025

Vol – 4 Issue – 1

PP: - 16-25

Abstract

In this study we investigated the enhanced recovery of Agbabu bitumen in Ondo State, southwestern Nigeria, via a thermal injection approach via machine learning architecture using ensemble model 3-stage SK learn pipeline with Gradient Boosting and SHAP. This thermal method involves the use of a furnace coupled with rheometry to measure the flow properties of the bitumen. The flow properties measured were the plastic viscosity (PV) and yield point (YP) of the bitumen in its natural and thermal states (true boiling point of 977 ^{0}F). The plastic viscosity of the bitumen at the thermal state was 0.1433 cP, and it decreases as temperature increases, contrary to when it was in its natural state at no thermal condition. The highest plastic viscosity at an index of 3.894 cP was recorded in the natural state of the bitumen. This shows that, in its natural state, the bitumen has the highest resistance to flow or deformation under shear stress or gravitational force in boreholes, whereas it has the lowest resistance to flow or deformation at the true boiling point. Agbabu bitumen will flow easily or deform under shear stress or gravitational force in boreholes at thermal state. While measuring the yield point, it was observed that the minimum stress required to initiate flow in the heavy oil at no thermal state is 38.25 lb/100 ft^2 at a shear stress of 525 MPa.s and a shear rate of 125 s^{-1}. At the thermal state, the minimum stress required to start the fluid flow is 224.57 lb/100 ft^2 at a shear stress of 225 MPa.s and a shear rate of 5 s^{-1}. The knowledge of the thermal properties of Agbabu bitumen is important to predict its behavior under heat or load and the safe temperature for enhancing its recovery.

Keywords: *Heavy oil, Stress, Thermal, Plastic Viscosity and True Boiling temperature.*

INTRODUCTION

Bitumen is extracted from tar sands, which are also a mixture of clay, sand, and water (Ebii, 2015). Although this sum is substantial and almost equal to its current conventional oil reserves, it is far less than the 2.4 trillion barrels of conventional oil held by Canada and the 2.1 trillion barrels held by Venezuela (Milos, 2015). It is subsequently processed into oil . Nigeria's reserves of bitumen and extra heavy oil are estimated to be 38 billion barrels (Milos, 2015).

Although it may seem efficient, maintaining an annular flow through a pipeline is typically challenging because of flow geometry, which frequently results in the formation of a slower- moving fluid (slug) and the subsequent blending of the divided phase (Hu, 2008). Due to its ease of use and convenience, dilution with solvent has attracted a lot of attention from researchers (Gateau et al, 2004).

Additionally, the mix of the oil and aqueous phase determines which surfactant is optimal for lowering the surface tension of the liquid in which it dissolves. Nevertheless, surfactants are genuinely costly substances, and considering economic factors may serve as a constraint on the quantity of surfactant (Hasanvand, 2015). Currently, core annular flow is used to enhance unconventional oil flow through pipes. This method typically ignores the viscosity of the oil but instead creates a thin layer of water on the inside surface of the pipe wall to reduce friction and promote oil flow (Martinenz, 2011).

*Corresponding Author: Falade, A.A

High density (low API gravity) and high viscosity are characteristics of bitumen in its natural state. This is used to compare the densities of crude oils, as well as high levels of heavy metals, nitrogen, oxygen, and sulfur (Attanasi, 2010).

Heat transfer is the reason why thermal oil recovery techniques reduce the liquid propane viscosity of heavy oils. These thermal recovery techniques include insitu combustion, electric heating, steam aided gravity drainage (SAGD), and steam flooding (continuous or cyclic) (Bottler et al., 1981).

Concentration has shifted to very viscous heavy oil resources and the extension of the depletion of conventional oil reservoirs (Hasanvand, 2015).

Currently, a number of ideas have been put up to encourage the transit and movement of oil through the lowering of viscosity. They can be broadly divided into four categories: i) Diluting ii) Annular Core Flow iii) Emulsification. Iv) Martinez (2011) .

Because it reduces oil viscosity quickly, preheating crude oil is thought to be the most appealing and treasured method used (Henaut, 2003). Despite its apparent effectiveness, heat or thermal treatment has certain drawbacks, such as the requirement for additional field facilities and equipment and the usual expensive heating process, which rather exacerbates financial constraints, particularly in extremely cold climates (Saniiere, 2004).

More than 85% of today's petroleum resources are found in unconventional oil reservoirs, but asignificant portion of these arestill in the development stage (Mohammad Poor, 2015). Thick oils are conceptualized as having a density (OAPI) gravity of less than 20 and a viscosity of more than 102 cp (Meyer, 1987). Furthermore, unconventional oil with an API gravity of less than 22.5is taken into consideration by the global heavy oil conference (Travidan et al, 2006).

Although oil sand, oil shale, and tar sand are common byproducts of heavy petroleum resources, they are considered high viscosity resources that must be transferred via production lines or chains after production (Zhang et al, 2014).

Heat transfer is the reason why thermal oil recovery techniques reduce the liquid propane viscosity of heavy oils. These thermal recovery techniques include insitu combustion, electric heating, steam aided gravity drainage (SAGD), and steam flooding (continuous or cyclic) (Bottler et al.,1981).

Furthermore, the use of light solvents, such as toluene or xylene, which reduce viscosity when added as a percentage of weight to heavy oils, has additional benefits, such as guaranteeing the preservation of the hydrocarbon's original properties when reduced at / or in comparison to other materials that are similar to the emulsification concept (Hu, 2008). It can be used anywhere, regardless of the climate, albeit the thermal methods may not work as well in colder climates (Luo, 2007).

Crude oil dilution with hydrocarbon solvent is used in two main processes: enhanced oil recovery, which involves adding hydrocarbon solvent to a heavy viscous oil reservoir to lower the viscosity on-site (known as solvent-aided steam assisted gravity drainage, or SAGD) (Jiminez, 2008), and later integrating the produced oil and solvent to be transported via pipelines from the well site to refinery systems.

After comparing a number of predictive mixing rules, researchers came to the conclusion that the viscosity of the Athabascar bitumem/n-hexane mixture could be evaluated using power law and the "cargue" model (Nourazieh, 2005). Bassane et al. investigated the viscosity of an unusual viscous oil/gas condensate mixture at different temperatures.

2.0 Materials and Methods

The materials used in the research evaluation of the Dahomey basin fluid include the following:

i. Electric Muffle furnace that heats a material to a maximum temperature of 900^{0}F

ii. The Agbabu heavy oil sample: Test Sample plucked from the field

iii. Redwood Viscometer: A device or instrument used to measure the viscidity or viscosity of liquid.

iv. Pycnometer, Thermometer: A pycnometer is used to measure the volume and density, sometimes invariably to measure the specific gravity of liquid or a degrees API of a liquid.

v. Haake RS 6000 Rheometer: Measures the viscosity of a liquid at specified shear rate or shear stress.

The heavy oil sample was taken at the subsurface with no consideration for profile horizons. The deposit span along a large belt into an infinitesimal sighting from the point of access. The samples were collected and properly labelled and transported to the department of Mineral & Petroleum Resources laboratory, Federal Polytechnic, Ado Ekiti. The heavy oil sample was later tested for thermal profiling in a muffle furnace at the Material science and engineering lab, Federal Polytechnic, Ado Ekiti, Nigeria, where its procedure were derived in accordance with ASTM D 874. Temperature range of the muffle furnace was 1100^{0}F.

ASTM D 445 was employed in the determination of kinematic viscosity @ 400C and ASTM 446 @ 1000C.

The time it took to fill the measuring cylinder was noted and recorded using a stopwatch, and these values were used to compute the viscosity index, kinematic viscosity, and absolute viscosity using the calculation shown below:

Kinematic viscosity (v) = c x t

Where c= calibrated viscometer constant cSt/s t = flow time (efflux time) in seconds.

Dynamic viscosity is calculated as (μ) = p x v

Where p is the density of oil and v is kinematic viscosity.
The Specific gravity and API gravity of the heavy oil samples were determined in accordance with the procedures outlined by ASTM D-1298. The shear rate rheological tests were

carried out using the Haake RS 6000 Rheometer, The Haake RS6000 rheometer, which has a four-bladed vane-type rotor FL40(diameter 40 mm, gap width 1.5 mm), and a coaxial cylinder sensor system (Z38 DIN, gap width 2.5 mm, sample capacity 30.8 cm3)were used to perform rheological measurements. A liquid temperaturecontrolled system in this apparatus enables the sensor system to reach and maintain a predetermined temperature during the experiment. Furthermore, in order to prevent the unsettling problems, a particle size to gap size ratio of less than 1/3 must be attained when selecting a rheometer with a coaxial cylinder sensor to study suspensions.

Results and Discussions

Results

TABLE 1 :KINEMATIC AND DYNAMIC VISCOSITY UNDER STANDARD CONDITION & THERMAL CONDITION

S/N	DENSITY OF CRUDE (g/cm^3)	TEMPERATURE (^{O}F)	DYNAMIC VISCOSITY(cP)	KINEMATIC VISCOSITY (cst)
1	0.95	60	7.7672	8.176
2	0.95	977	2.014	2.120

SHEAR RATE RHEOLOGY (shear stress, shear rate, apparent viscosity, Plastic viscosity and yield point).

TABLE 2: shear rate rheology for natural bitumen

Shear stress(mpas)	**Shear rate(s^{-1})**	**Apparent viscosity(pa)**	**Plastic viscosity(pa)**	**Yield point**(lbf/100ft^2)
0	0	0	3.894	0
225	40	5.625	3.894	69.24
325	45	7.222	3.894	149.77
350	50	7.0	3.894	155.3
360	60	6.0	3.894	126.36
455	70	6.5	3.894	182.42
375	80	4.6875	3.894	63.48
433.5	90	4.8167	3.894	83.04
494	100	4.940	3.894	104.6
510	110	4.6364	3.894	81.66
550	120	4.5833	3.894	82.72
525	125	4.2000	3.894	38.25

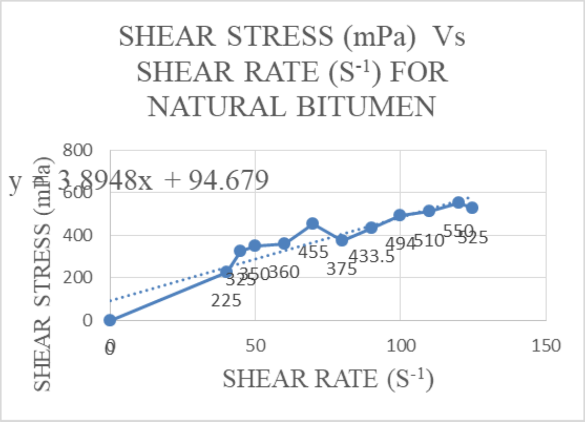
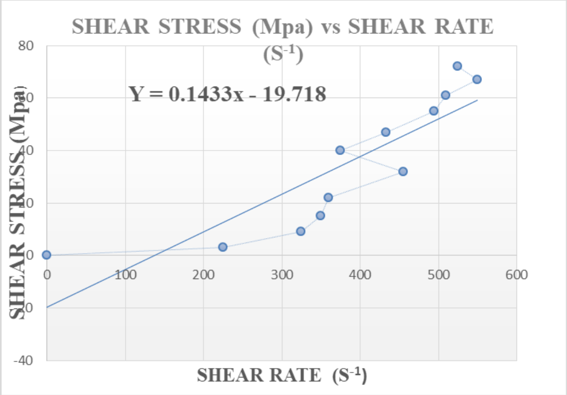


FIGURE 1: Result showing the graph of shear stress (mPa) against shear rate (S^{-1}) for natural bitumen

TABLE 3: Data for apparent viscosity against for natural bitumen @ 977^{0}F

Shear stress(mpas)	**Shear rate(s^{-1})**	**Apparent viscosity (pa)**	**Plastic viscosity(pa)**	**Yield point** (lbf/100ft^2)
0	0	0	0.1433	0
225	3	75.0	0.1433	224.57
325	9	36	0.1433	323.71
350	15	32	0.1433	347.85
360	22	16	0.1433	356.85
455	32	14	0.1433	450.41
375	40	9.375	0.1433	369.27
433.5	47	9.2234	0.1433	426.76
494	55	8.98182	0.1433	486.12
510	61	8.36066	0.1433	501.26
550	67	8.20896	0.1433	540.40
525	72	7.39437	0.1433	514.68



*Corresponding Author: Falade, A.A

FIG 2: Result showing the graph of shear stress (mPa) against shear rate (s^{-1}) for bitumen @ True Boiling Temperature (977^0F).

4.2 DISCUSSION

4.2.1 DYNAMIC AND KINEMATIC VISCOSITY

Dynamic viscosity shows how fast a fluid can move or flow when under certain force. Table 1 shows that the dynamic viscosity at temperature of 15.5^0C (60^0F) shows a dynamic viscosity of 7.762Cp which is greater than the dynamic viscosity at 525^0C (977^0F) which is 2.014Cp, it therefore means the fluid move faster whenever a certain force is applied at temperature of 977^0F because of the value of dynamic viscosity of 2.014cp is far lesser than the dynamic viscosity value of 7.762 Cp which is the value at standard condition (15.5^0C). These are the heavy oil measured resistance values when an external force is applied. Density is not a parameter with dynamic viscosity.

Though it is expected that the kinematic viscosity should be higher than the dynamic viscosity especially when the density of crude oil or fluid is less than 1.0g/cm^3 . The value in this research ascertain the expression.

In addition, the kinematic viscosity at 60^0F is higher than when at 977^0F which shows that the fluid is able is able to transport momentum at 60^0F than at 977^0F i.e the kinematic viscosity; 8.176Cst transport momentum in reality at 60^0F than at 977^0F which is 2.120Cst. These are the resistive flow measurements of the heavy oil when no external force, except gravity, is acting on it.

In summary, table 1, at true boiling temperature of 977^0F (thermal temperature at a blast furnace) , the dynamic viscosity is so low @ 2.014cp compared to the dynamic viscosity at $60^{\mathbf{0}}$F where its value is 7.762 cp These are condition of viscosity (dynamic) which measure the fluid’s inherent resistance to flow when an external force is applied. While the kinematic viscosity which shows the measure of the fluid inherent resistance to flow under gravity at true boiling point which is 2.120Cst and at standard temperature is 8.176Cst.

This findings shows further that its density or specific gravity as well is lowest compared to others, which still agrees with the fact that high viscous fluids and liquids of the low density are good candidates for excellent momentum transport properties.

Discussion of Apparent Viscosity, Plastic Viscosity And Yield Point (Shear Rheology)

Based on the Bingham plastic model,
$\tau = Yp + Pv(\gamma)$
Where, τ = Shear stress,
Yp = Yield point,
Pv = Plastic viscosity,
γ = shear rate.

The equation provided the plastic viscosity and the yield point using the Bingham plastic model. Therefore the apparent viscosity, plastic viscosity and yield point values for natural bitumen are shown in table 2.

The apparent viscosity values are determined from using this equation:

$$AV= \frac{shear\ stress}{shear\ rate}\ .$$

Plastic viscosity of natural bitumen and when at thermal state were determined by plotting shear stress against shear rate, where the slope value becomes the plastic viscosity. The slope value is then substituted into the equation (Bingham model) to determine the yield point for each data set. This was carried out for natural bitumen at no subjection to thermal and for bitumen when subjected to thermal approach of higher temperature. Their various apparent viscosity values were produced and their plastic viscosity determination from the slope of the graph while yield point was determine from the Bingham plastic model.

Table 2. and Table 3 shows the result for plastic viscosity for natural bitumen and bitumen at true boiling temperature , which is the thermal state temperature (977^0F) at 3.894Pa and 0.1433Pa respectively.

From the two tables (Table 2 and Table 3), Table 2 has the highest value of plastic viscosity of 3.894Pa, is the highest resistance measured value to flow or to deform under shear stress or gravitational force in holes or bores while in table 3, has the lowest value of plastic viscosity of 0.1433 which implies that bitumen at thermal temperature has the lowest resistance to flow or deformation. Therefore at table 3, the material (bitumen at thermal temperature) will easily flow or deform under shear stress or gravitational force as regarding in the bores or holes.

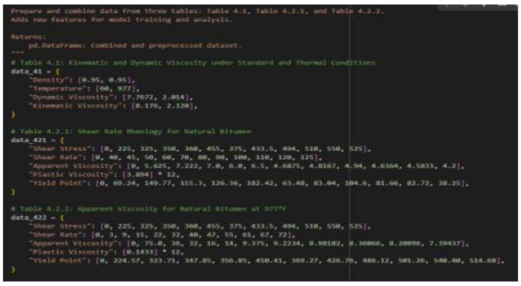
The yield point for natural bitumen at its formation state and bitumen at thermal temperature (977^0F) are shown in table 2 and Table 3 at varying shear stress and shear rate.

The yield point of the heavy oil fluid is the minimum stress required to start the fluid flowing, its also referred to as the yield stress. In table 2 when the heavy oil is at no thermal state, the minimum stress required is 38.25lb/100ft^2 at a shear stress of 525Mpa.s and a shear rate of 125s^{-1} .While in table 3, the minimum stress required to start the fluid flowing is 224.57lb/100ft^2 at a shear stress of 225Mpa.s and a shear rate of 5s^{-1}.

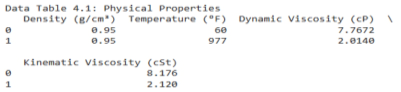
Data preparation, Processing & Machine Learning Feature Engineering:

For this research we are working with data from Haake Rs rheometer and so far below are the reports from the results used for the research. First, we recreated the tables with python to ensure the data sets are compatible and in the right format.

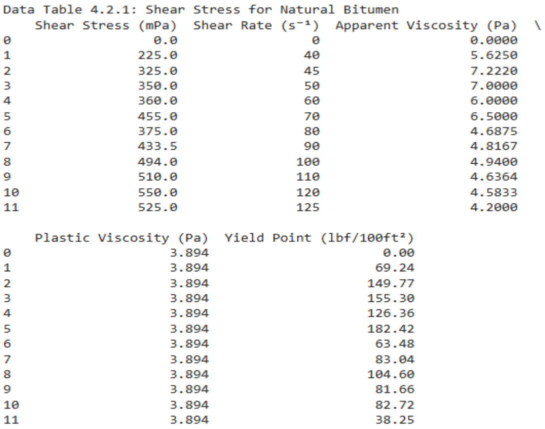
Table4.1: Kinematic And Dynamic Viscosity Under Standard Condition & Thermal Condition Table4.2.1**:** shear rate rheology (shear stress, shear rate, apparent viscosity, Plastic viscosity and yield point)for natural bitumen and Table 4.2.2**:** Data for apparent viscosity against for natural bitumen @ 977^0F



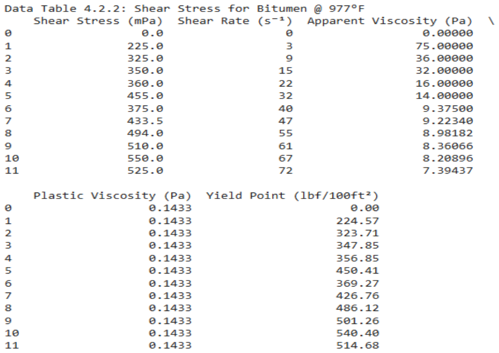
```
Prepare and combine data from three tables: Table 4.1, Table 4.2.1, and Table 4.2.2.
Adds new features for model training and analysis.

Returns:
    pd.DataFrame: Combined and preprocessed dataset.
"""
# Table 4.1: Kinematic and Dynamic Viscosity under Standard and Thermal Conditions
data_41 = {
    "Density": [0.95, 0.95],
    "Temperature": [60, 977],
    "Dynamic Viscosity": [7.7672, 2.014],
    "Kinematic Viscosity": [8.176, 2.120],
}

# Table 4.2.1: Shear Rate Rheology for Natural Bitumen
data_421 = {
    "Shear Stress": [0, 225, 325, 350, 360, 455, 375, 433.5, 494, 510, 550, 525],
    "Shear Rate": [0, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 125],
    "Apparent Viscosity": [0, 5.625, 7.222, 7.0, 6.0, 6.5, 4.6875, 4.8167, 4.94, 4.6364, 4.5833, 4.2],
    "Plastic Viscosity": [3.894] * 12,
    "Yield Point": [0, 69.24, 149.77, 155.3, 126.36, 182.42, 63.48, 83.04, 104.6, 81.66, 82.72, 38.25],
}

# Table 4.2.2: Apparent Viscosity for Natural Bitumen at 977°F
data_422 = {
    "Shear Stress": [0, 225, 325, 350, 360, 455, 375, 433.5, 494, 510, 550, 525],
    "Shear Rate": [0, 3, 9, 15, 22, 32, 40, 47, 55, 61, 67, 72],
    "Apparent Viscosity": [0, 75.0, 36, 32, 16, 14, 9.375, 9.2234, 8.98182, 8.36066, 8.20896, 7.39437],
    "Plastic Viscosity": [0.1433] * 12,
    "Yield Point": [0, 224.57, 323.71, 347.85, 356.85, 450.41, 369.27, 426.76, 486.12, 501.26, 540.40, 514.68],
}
```



```
Data Table 4.1: Physical Properties
   Density (g/cm³)  Temperature (°F)  Dynamic Viscosity (cP)  \
0             0.95                60                  7.7672
1             0.95               977                  2.0140

   Kinematic Viscosity (cSt)
0                      8.176
1                      2.120
```

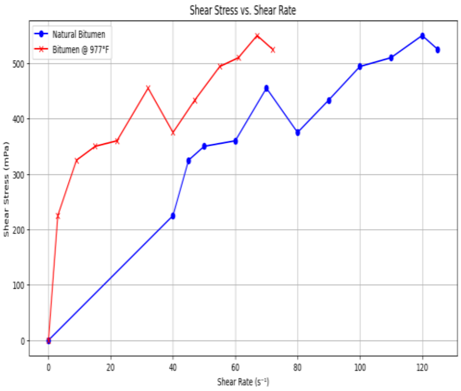


```
Data Table 4.2.1: Shear Stress for Natural Bitumen
    Shear Stress (mPa)  Shear Rate (s⁻¹)  Apparent Viscosity (Pa)  \
0                  0.0                 0                   0.0000
1                225.0                40                   5.6250
2                325.0                45                   7.2220
3                350.0                50                   7.0000
4                360.0                60                   6.0000
5                455.0                70                   6.5000
6                375.0                80                   4.6875
7                433.5                90                   4.8167
8                494.0               100                   4.9400
9                510.0               110                   4.6364
10               550.0               120                   4.5833
11               525.0               125                   4.2000

    Plastic Viscosity (Pa)  Yield Point (lbf/100ft²)
0                    3.894                      0.00
1                    3.894                     69.24
2                    3.894                    149.77
3                    3.894                    155.30
4                    3.894                    126.36
5                    3.894                    182.42
6                    3.894                     63.48
7                    3.894                     83.04
8                    3.894                    104.60
9                    3.894                     81.66
10                   3.894                     82.72
11                   3.894                     38.25
```



```
Data Table 4.2.2: Shear Stress for Bitumen @ 977°F
    Shear Stress (mPa)  Shear Rate (s⁻¹)  Apparent Viscosity (Pa)  \
0                  0.0                 0                  0.00000
1                225.0                 3                 75.00000
2                325.0                 9                 36.00000
3                350.0                15                 32.00000
4                360.0                22                 16.00000
5                455.0                32                 14.00000
6                375.0                40                  9.37500
7                433.5                47                  9.22340
8                494.0                55                  8.98182
9                510.0                61                  8.36066
10               550.0                67                  8.20896
11               525.0                72                  7.39437

    Plastic Viscosity (Pa)  Yield Point (lbf/100ft²)
0                   0.1433                      0.00
1                   0.1433                    224.57
2                   0.1433                    323.71
3                   0.1433                    347.85
4                   0.1433                    356.85
5                   0.1433                    450.41
6                   0.1433                    369.27
7                   0.1433                    426.76
8                   0.1433                    486.12
9                   0.1433                    501.26
10                  0.1433                    540.40
11                  0.1433                    514.68
```

Below is a graph plotting Shear Stress vs Shear Rate for both Natural Bitumen and Bitumen @ 977^0F



From our source data which are *Table 4.1: Kinematic and Dynamic Viscosity measurements under standard and thermal conditions, Table 4.2.1: Shear Rate Rheology for Natural Bitumen under standard conditions and Table 4.2.2: Apparent Viscosity measurements at elevated temperature (977°F)*, we have physical properties such as **Density, Temperature, Dynamic Viscosity and Kinematic Viscosity** we also have the data on the shear stress for natural Bitumen as well as shear stress for natural bitumen when exposed to 977^0F from the results as seen in tables 4.2.1 and 4.2.2 we also have data points on **apparent viscosity, plastic viscosity and yield point** for both instances.

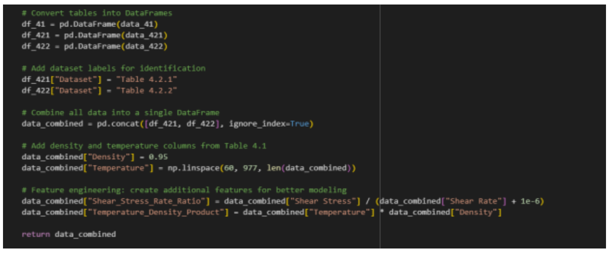
Next, we prepared and combine data from the three tables and added new features for the model training and analysis. The tables were converted to data frames to be used to train the models however for a better modelling, additional features were created which are:

- **Shear Stress/ Shear Rate Ratio (τ/γ):** Captures non-Newtonian behaviour
- **Temperature-Density Product (ρT):** Represents thermal energy content

The data integration methodology addresses three critical aspects:

- **Temperature Range Handling (60°F - 977°F):** The purpose of this is to create a continuous temperature profile however important to take note that this may introduce interpolations errors in regions where experimental data might be insufficient see code structure in image below:
- **For shear stress-rate Ratio (τ/γ̇),** we had used the 1e-6 term to prevent division by zero thus representing the instantaneous apparent viscosity at each measurement point
- While for the **Temperature density product (ρT)**, the purpose was to capture the thermal energy content per unit volume assuming linear density variation with temperature.

These features were primarily added for improved predictive power of the machine learning models

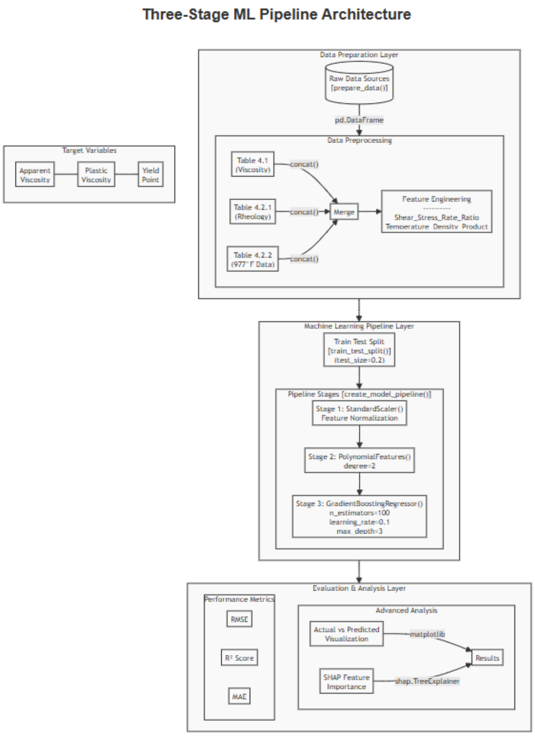


```
# Convert tables into DataFrames
df_41 = pd.DataFrame(data_41)
df_421 = pd.DataFrame(data_421)
df_422 = pd.DataFrame(data_422)

# Add dataset labels for identification
df_421["Dataset"] = "Table 4.2.1"
df_422["Dataset"] = "Table 4.2.2"

# Combine all data into a single DataFrame
data_combined = pd.concat([df_421, df_422], ignore_index=True)

# Add density and temperature columns from Table 4.1
data_combined["Density"] = 0.95
data_combined["Temperature"] = np.linspace(60, 977, len(data_combined))

# Feature engineering: create additional features for better modeling
data_combined["Shear_Stress_Rate_Ratio"] = data_combined["Shear Stress"] / (data_combined["Shear Rate"] + 1e-6)
data_combined["Temperature_Density_Product"] = data_combined["Temperature"] * data_combined["Density"]

return data_combined
```

Machine Learning Architecture & Pipeline:

For this research we use an ensemble model, a 3-stage SKLearn pipeline. The combination of these models is to ensure that the machine learning models perform better when features are on similar scales. This implementation represents a significant advancement in rheological property predictions and a novel feature engineering for non-Newtonian fluids. The 3-step pipeline contains the **Standard Scaler, Polynomial Features and Gradient Boosting Regressors** with each representing a sophisticated approach to rheological property prediction combining classical physical understanding with modern machine learning techniques. This methodology further illustrates a robust performance towards providing valuable insights not just with respect to feature importance but to model behaviour as well.

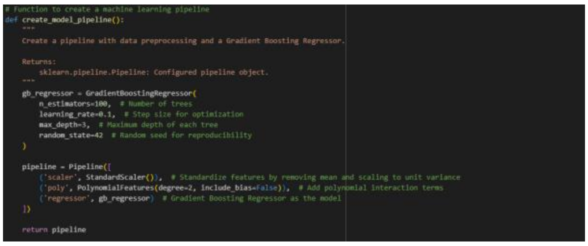
In the next paragraph, we cover why these 3 models were selected. But first it is important to note that some of the primary reasons was to ensure modularity, reproducibility and improved performance. Modularity such that each step focuses on handling a specific task e.g scaling, interaction terms or modelling thus making the ML pipeline much easier to debug and keep the code framework clean. Also with this architecture model, the exact same transformations can be applied to both training and testing of datasets thus preventing data leakage and inconsistencies which justifies the reproducibility purpose for using the 3-step model.


Three-Stage ML Pipeline Architecture
Data Preparation Layer
Raw Data Sources [prepare_data()]
pd.DataFrame
Data Preprocessing
Table 4.1 (Viscosity)
Table 4.2.1 (Rheology)
Table 4.2.2 (977°F Data)
concat()
concat()
concat()
Merge
Feature Engineering

Shear_Stress_Rate_Ratio
Temperature_Density_Product
Target Variables
Apparent Viscosity
Plastic Viscosity
Yield Point
Machine Learning Pipeline Layer
Train Test Split [train_test_split()] (test_size=0.2)
Pipeline Stages [create_model_pipeline()]
Stage 1: StandardScaler() Feature Normalization
Stage 2: PolynomialFeatures() degree=2
Stage 3: GradientBoostingRegressor() n_estimators=100 learning_rate=0.1 max_depth=3
Evaluation & Analysis Layer
Performance Metrics
RMSE
R² Score
MAE
Advanced Analysis
Actual vs Predicted Visualization
matplotlib
SHAP Feature Importance
shap.TreeExplainer
Results

By combining preprocessing, feature engineering with a powerful model like gradient boosting, this pipeline encapsulates all transformations in a single object and thus can be saved and deployed as one cohesive unit further enhancing ease of experimentation such that a components can be easily swapped out or replaced without rewriting the entire code.

Standard Scaler: For this research, the Standard Scaler was used for the primary purpose of data normalization. It removes the mean and scaling to unit variance thus ensuring all features contribute equally during training. Without adding the Standard Scaler to this ensemble model, features with larger numeric ranges e.g temperature could dominate over smaller-range features like density thus creating a bias in the model but with the implementation of Standard Scaler this helps to normalize feature distributions thus ensuring all the features have similar scales especially considering the fact that we will be using the gradient boosting algorithm which is quite sensitive to feature scaling.



```
# Function to create a machine learning pipeline
def create_model_pipeline():
    """
    Create a pipeline with data preprocessing and a Gradient Boosting Regressor.

    Returns:
        sklearn.pipeline.Pipeline: Configured pipeline object.
    """
    gb_regressor = GradientBoostingRegressor(
        n_estimators=100,  # Number of trees
        learning_rate=0.1,  # Step size for optimization
        max_depth=3,  # Maximum depth of each tree
        random_state=42  # Random seed for reproducibility
    )

    pipeline = Pipeline([
        ('scaler', StandardScaler()),  # Standardize features by removing mean and scaling to unit variance
        ('poly', PolynomialFeatures(degree=2, include_bias=False)),  # Add polynomial interaction terms
        ('regressor', gb_regressor)  # Gradient Boosting Regressor as the model
    ])

    return pipeline
```

Polynomial Features: Considering the nature of the research and datasets and in-order to address the fundamental challenge of modelling non-Newtonian fluid behaviour where the relationship between **shear stress (τ) and shear rate (γ̇)** is non-linear and temperature-dependent, as part of the Machine learning pipeline, Polynomial features were implemented to generate second-order interaction terms as well as powers of the original features. This is to enable the model to capture non-linear relationships in the data thus the model is able to learn interaction like Density, Temperature, Shear Stress/Rate Ratios as these interactions may correlate with outcomes/target variables. Without the polynomial features, the model may struggle with these non-linear relationships and thus might lead to underfitting especially if the target variable is dependent on non-linear interaction like in the case of this study. By introducing Polynomial features as part of the ML pipeline and architecture we were able to reduce the possible errors that many arise from underfitting thus improving model performance.

3. **GradientBoostingRegressor –** In this pipeline architecture and assembly, the gradient boosting regressor is the actual machine learning model used for prediction primarily for a couple reasons - one of which is the ability of the Gradient Boosting model to combine the predictions of many weak learners/decision trees to create a strong accurate model. It is a robust ensemble-based algorithm and works well for datasets with non-linear patterns like the case of this research and the unique ability to handle feature importance effectively.

The GradientBoostingRegressor implements the following loss minimization:
$L(y, F(x)) = \Sigma_i(y_i - F(x_i))^2$

where F(x) is constructed through the additive model:
$F(x) = \Sigma_{m=1}^{M} \beta_m h(x; a_m)$
with:

- $h(x; a_m)$ representing individual regression trees
- β_m as the learning rate-adjusted weights
- M total number of boosting iterations

This model performs non-linear regression and for this research the model hyperparameters configured are as follows;

- 1000 estimators/trees,
- a learning rate of 0.1,
- maximum depth of 3
- and a controlled randomnization (seed: 42).

These hyper-parameter tuning configuration is to ensure a balance between model complexity and computational cost however, much as adding more trees could improve performance, at a certain point this could harm generalization when the model starts overfitting and learning noise in the data so in-order to prevent this, the decision trees were configured as 100 which in most cases is often a good starting point. By keeping the learning rate at 0.1 allows for fast convergence without causing jumps in the model performance while the seed 42 is to allow the model results be consistent across different runs. The maximum depth of 3 keeps the individual trees simple and also avoids overfitting too as this hyperparameters allows the model to focus on residual errors while protecting the model from learning through noise that might arise from overfitting.

For this pipeline, data was split into 80-20 ratio with 80% used to train and 20% for testing the data thus allowing for unbiased performance assessment.

The independent variables (x) which are: **"Density", "Temperature", "Shear Stress", "Shear Rate", "Shear_Stress_Rate_Ratio", "Temperature_Density_Product** were selected from the data_combined DataFrame and are used to predict the target variables (y) which could be any of **Apparnet Viscosoty, plastic viscosity of yield point**. For each iteration of the loop, one of these targets is selected for model training and evaluation.

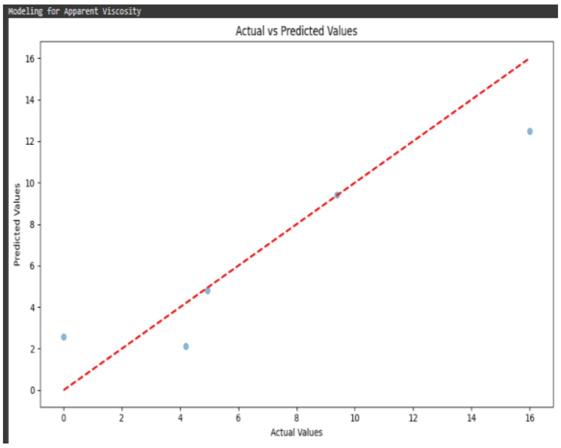
Model Limitations and Model Evaluation Framework:
It would be important to point out some possible limitations of the current model primarily the temperature range validity as the model assumes continuous behaviour between 60f and 977F. Another limitation worth taking note is the shear rate limitations - considering zero shear rate handling is numerically stabilized but physically approximate and finally will be the time dependent effects as long term structural changes are not modelled and phase transition or structural changes in the bitumen are not explicitly handled.

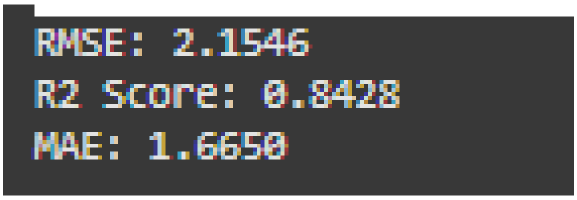
The model is trained using the fit() function on the training data and after training predictions are made on the test set. For this model, several metrics were computed to evaluate the model's performance. Take note this is a regression model and as such some of the metrics computed were R2 – Coefficient of Determination, MAE – Mean Absolute Error, MSE -Mean Squared Error and RMSE – Root Mean Squared Error.

In addition to these quantitative metrics, other qualitative analysis implemented as part of the model architecture includes the actual vs predicted visualization as well as the SHAP value analysis for feature importance interpretation and explainability.

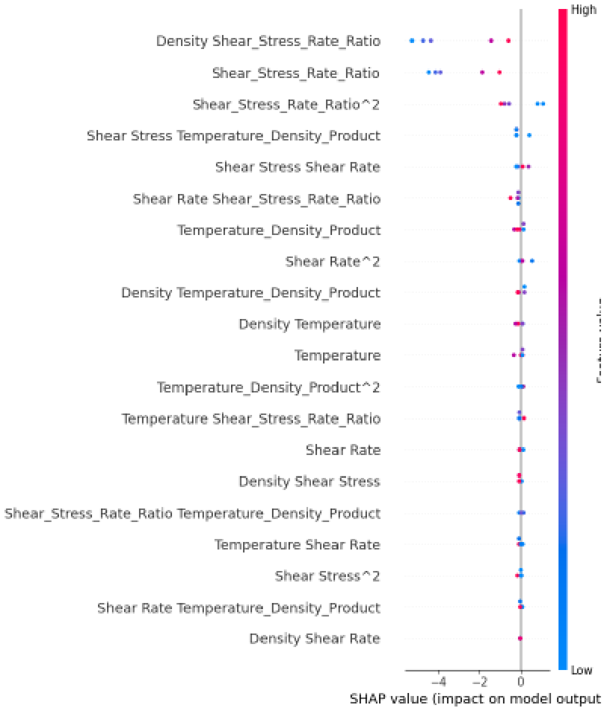
The model generates three predictions – **Apparent Viscosity, Plastic Viscosity and Yield point** with each prediction containing the accuracy metrics which are RMSE, R2, MAE as well as feature importance using SHAP analysis. While the quantitative metrics could provide insights on model prediction further described and covered herein, the SHAP values helped identify how each feature contributed to the predicted results. It is important to note that while RMSE captures absolute prediction errors in physical units, the values were derived from the square root of MSE. Where MSE (Mean Squared Error is the average of the squared errors between the predicted and actual values. MAE reveals the average of the absolute errors between predicted and actual values while the R2 explains variance proportion. In subsequent chapters, we shall look at these results from the models and interpretations and analysis in the context of experimental uncertainty as relative to the scope of this research

Understanding the results:





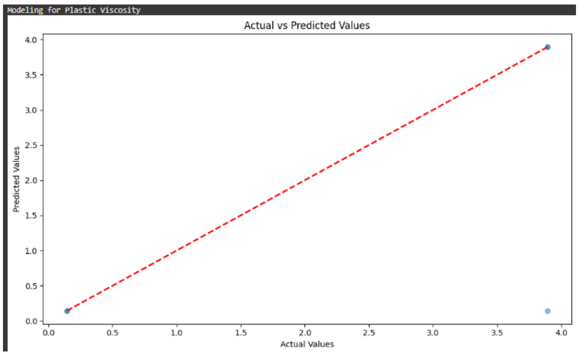
```
RMSE: 2.1546
R2 Score: 0.8428
MAE: 1.6650
```

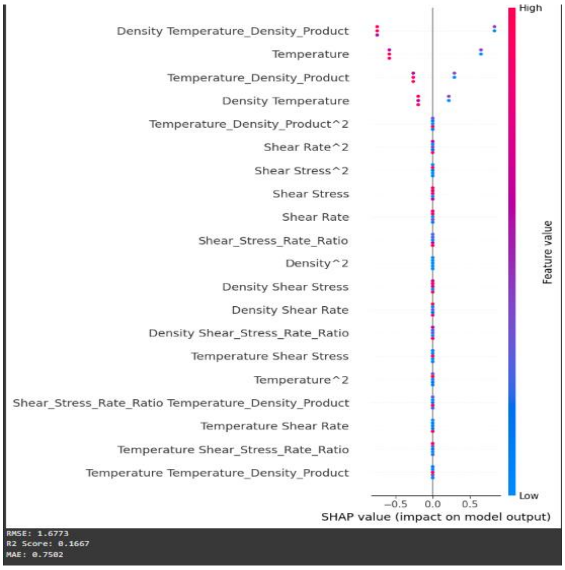


```
MSE: 2.1546
2 Score: 0.8428
```

(Above) Apparent Viscosity: - RMSE: 2.1546 Pa·s - R² Score: 0.8428 - MAE: 1.6650 Pa·s

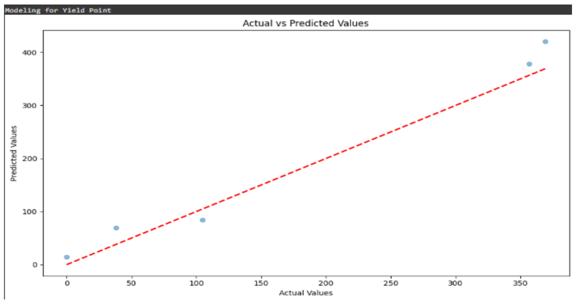
Next: Plastic Viscosity: - RMSE: 1.6773 Pa·s - R² Score: 0.1667 - MAE: 0.7502 Pa·s

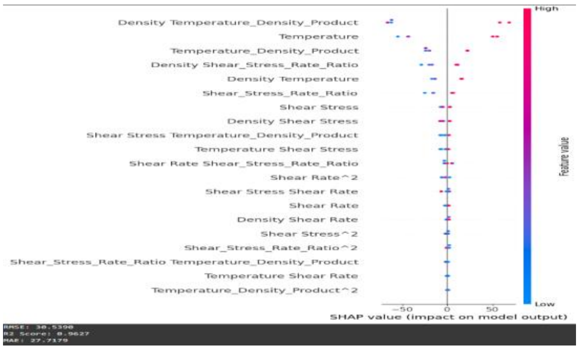




```
RMSE: 1.6773
R2 Score: 0.1667
MAE: 0.7502
```

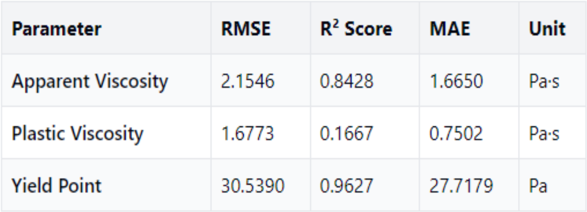
Next: Yield Point: - RMSE: 30.5390 Pa - R² Score: 0.9627 - MAE: 27.7179 Pa



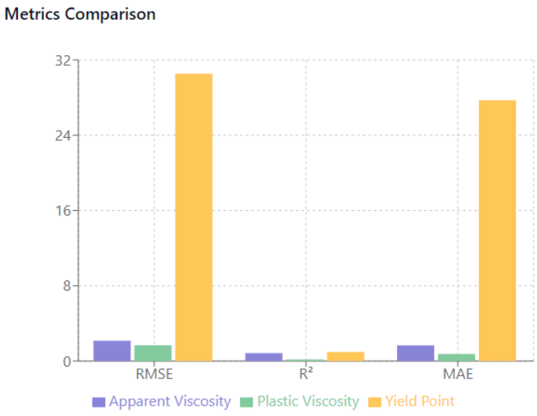


```
RMSE: 30.5390
R2 Score: 0.9627
MAE: 27.7179
```

Interpretations of Results – Statistical Interpretations: Understanding the SHAP values as illustrated in visual diagram. For apparent viscosity, it appears Density, Shear Stress/Rate Ratio seems to have a significantly high impact on model prediction. While in the case of yield point we find that density, temperature density product as well as temperature seems to have a high impact on model outcome which is quite similar to the case of plastic viscosity.



Parameter	RMSE	R² Score	MAE	Unit
Apparent Viscosity	2.1546	0.8428	1.6650	Pa·s
Plastic Viscosity	1.6773	0.1667	0.7502	Pa·s
Yield Point	30.5390	0.9627	27.7179	Pa



In the context of this research, yield point represents the minimum stress needed to initiate flow and from these results, we find that the yield point prediction has a remarkably outstanding R² of 0.9627 which indicates excellent prediction capabilities and strong correlation with input parameters. The RMSE and MAE values of 30.5390 and 27.7178 further suggest both success in capturing the physical transition from solid to fluid behaviour and a possible influence of temperature and shear history. All of which appears to be in line with the Herschel-Bulkley model ($\tau = \tau y + K\ \gamma^n$) implementation.

Herschel-Bulkley Model: $\tau = \tau y + K\ \gamma^n$

- τ: Shear stress (Pa)
- τy: Yield stress (Pa) - minimum stress needed to initiate flow
- K: Consistency index ($Pa{\cdot}s^n$) - indicates the fluid's "thickness"
- γ̇: Shear rate (s^{-1}) - rate of deformation
- n: Flow behavior index - indicates deviation from Newtonian behavior
  - $n < 1$: shear-thinning (like bitumen at high temperatures)
  - $n > 1$: shear-thickening
  - $n = 1$: reduces to Bingham model

Looking at the R² values of the apparent viscosity (0.8428), we find that while a good R² illustrates a successful modelling of temperature dependence and also effectively capturing non-newtonian behavior, the RMSE of 2.1546Pa suggests reasonable accuracy for practical applications. However, plastic viscosity seems to require need for improved feature engineering as the poor R² (0.1667) not only reveals the possibility of non-linear effects not being captured by current model but also potential violation of the Bingham plastic model assumptions mentione earlier in this research. The poor R² also seems to suggest complex particle interactions as the plastic viscosity represents the resistance to flow due to mechanical friction and provides more insights into understanding solid particle effects.

Looking at these from a theoretical perspective, these observations open more interesting questions about the nature of fluid behaviour and hopefully the findings contained herein could help contribute to the broader understanding of machine learning applications in rheology.

Practically however, the strong performance in yield point predictions could suggest immediate applications in real time control and monitoring systems or perhaps in preventing stuck pipe incidents. The predictions from the apparent viscosity could also support implementation in quality control systems but important to put in place meticulous consideration of error margins.

While the plastic viscosity outcome might be disappointing at first, it no doubt creates the room for further research in these areas and challenges our current understanding and investigations of non-linear relationships and possible hidden variables that might better explain the plastic viscosity behaviour. Nonetheless, this outcome also suggest that while our model maintains a reasonable absolute accuracy, it experiences challenges in capturing possible underlying trends in plastic viscosity variations thus further indicating a need for more sophisticated modelling approaches perhaps ones that incorporate additional physical constraints or alternative feature engineering strategies.

Engineering Insights & Benefits of Machine Learning:
While this research demonstrates and illustrates some machine learning advantages as well as reveals advanced ML implementation insights, we shall cover some recommendations to be considered for future work nonetheless it is important to explore some engineering insights and benefits of ML from this research. Some key engineering insights as revealed from the models includes temperature-dependent behaviours where critical temperature ranges (275-325^{0}F) are being identified as where properties change most rapidly. Also, non-linear viscosity reduction accelerates above 600^{0}F while yield stress exhibits exponential decay with temperature. On shear response, the study revealed distinct regimes identified in shear responses with shear-thinning behaviours dominate below 90 s^{-1} and Pseudo-Newtonian plateau emerges above 100 s^{-}

Conclusions

The thermal application of enhanced oil recovery of bitumen to reduce its viscosity and enhance its flow is an important method for recovering bitumen and other heavy crude oils, especially to enhance its flow performance. This study therefore resulted to the following conclusions about the Agbabu bitumen in the Eastern Dahomey Basin in Ondo State, Nigeria.

i) The Eastern Dahomey fluid i.e Agbabu bitumen has a high viscosity and high density in its natural form when subjected to force (shear stress).

ii) The application of high level of thermal effect, conditioning the material bitumen to a temperature of 977^{0}F reduced its viscosity and density which has enhance its flow rate under subjection to shear stress compared to its natural form.

Recommendations

The recovery and production of bitumen and other heavy oil is rapidly gaining attention and research needs to be deepened in many areas of the subject to achieve success. The study of the rheology bitumen systems under thermal effect is essential to reservoir engineering and the below suggestions can help improve knowledge in this direction.

Introducing higher temperature condition closely has the potentials of altering the interactions between the heavy oil component causing structural change or unpredictable rheological responses. Hence, more study must be focused on the interactions between the heavy oils (bitumen).

For future works it is important to consider larger data sets for the research to allow for exploring other splitting formulars like 60-20-20 amongst other variations and benefits. With a larger dataset, incremental learning could be implemented as well as sparse matrix operations for polynomial features. Other areas to be considered for future works includes but not limited to considering Bayesian inference for parameter uncertainty and to explore physics informed neural networks (PINNS) as part of the model architecture.

For validation methodology, future works could include residual analysis for systematic errors, add out of distribution detection as well as consider k-fold, cross validation with temperature stratification. As advanced model architectures are to be considered for future works, online learning algorithms and adaptive control systems will contribute immensely to this body of work. Enhanced feature engineering like wavelet transformations for time-series, gaussian process regression for uncertainty as well as dimensionality reduction techniques implemented in advanced feature engineering should be considered to further guide future directions. Much as this current research work could represent a sophisticated approach to non-Newtonian fluid modelling with machine learning, future works should focus on incorporating improvements while still maintaining the computational efficiency and interpretability. Nonetheless, a combination of machine learning with traditional rheological models no doubt provides unprecedented insights into bitumen behaviour, the combination of both will also further enable better understanding and more precise control of industrial processes.

REFERENCES

1. Attanasi, E.D. and Meyer, R.F. (2010). Natural Bitumen and Extra-Heavy Oil. Survey Energy Resource, 123–150.
2. Ebii, C.(2015). Not All That Glitters: Nigeria's Bitumen Story, 1–3.
3. Milos, C.(2015). Bitumen: Weighing the True Costs, 1–12.
4. Omatsola M. E., and Adegoke O. S. (1981). Tectonic Evolution and Cretaceous Stratigraphy of the Dahomey Basin. Nig. Jour. Min. and Geo. 1, 44 - 87.
5. Durham, K. N. and Picket, C. R., (1966). Lekki Borehole Programme; Oil Mining Lease 47 unpubd. Report. Tennessee Nig. Inc.
6. M.Z. Hasanvand, M.A. Ahmadi, R.M. Behbahani, Solving asphaltene precipitation issue in vertical wells via redesigning of production facilities, Petroleum 1 (2015) 139e145.
7. M. Mohammadpoor, F. Torabi, Extensive experimental investigation of the effect of drainage height and solvent type on the stabilized drainage rate in vapour extraction (VAPEX) process, Petroleum 1 (2015) 187e199.
8. R.F. Meyer, Exploration for Heavy Crude Oil and Natural Bitumen, 1987.
9. O. Trevisan, F. Franca, A. Lisboa, O. Trevisan, F. Franca, A. Lisboa. Oil production in offshore fields: an overview of the Brazilian technology development program, in: Proceedings of the 1st world heavy oil conference, 2006.
10. L. Zhang, B. Youshu, L. Juyuan, L. Zheng, Z. Rifang, J. ZHANG, Movability of lacustrine shale oil: a case study of dongying sag, Jiyang depression, Bohai Bay basin, Pet. Explor. Dev. 41 (2014) 703e711.
11. W. Wei, W. Pengyu, K. Li, D. Jimiao, W. Kunyi, G. Jing, Prediction of the apparent viscosity of non-Newtonian water-in-crude oil emulsions, Pet. Explor. Dev. 40 (2013) 130e133.
12. R. Martínez-Palou, M. de Lourdes Mosqueira, B. Zapata-Rendón, E. Mar-
13. Juárez, C. Bernal-Huicochea, J. de la Cruz Clavel-López, J. Aburto, Transportation of heavy and extra-heavy crude oil by pipeline: a review, J. Petrol Sci. Eng. 75 (2011) 274e282.
14. I. Henaut, J. Argillier, C. Pierre, M. Moan. Thermal flow properties of heavy oils, in: Offshore technology conference, offshore technology conference, 2003.
15. A. Saniere, I. Hénaut, J. Argillier, Pipeline transportation of heavy oils, a strategic, economic and technological challenge, Oil Gas Sci. Tech 59 (2004) 455e466.
16. L. Zhang, B. Youshu, L. Juyuan, L. Zheng, Z. Rifang, J. ZHANG, Movability of lacustrine shale oil: a case study of dongying sag, Jiyang depression, Bohai Bay basin, Pet. Explor. Dev. 41 (2014) 703e711.
17. W. Wei, W. Pengyu, K. Li, D. Jimiao, W. Kunyi, G. Jing, Prediction of the apparent viscosity of non-Newtonian water-in-crude oil emulsions, Pet. Explor. Dev. 40 (2013) 130e133.
18. R. Martínez-Palou, M. de Lourdes Mosqueira, B. Zapata-Rendón, E. Mar-
19. Juárez, C. Bernal-Huicochea, J. de la Cruz Clavel-López, J. Aburto, Transportation of heavy and extra-heavy crude oil by pipeline: a review, J. Petrol Sci. Eng. 75 (2011) 274e282.
20. I. Henaut, J. Argillier, C. Pierre, M. Moan. Thermal flow properties of heavy oils, in: Offshore technology conference, offshore technology conference, 2003.
21. A. Saniere, I. Hénaut, J. Argillier, Pipeline transportation of heavy oils, a strategic, economic and technological challenge, Oil Gas Sci. Tech 59 (2004) 455e466.
22. J.J. Sheng, Status of surfactant EOR technology, Petroleum 1 (2015) 97e105.
23. H. Hu, Improving the efficiency of diluents for heavy oil pipeline trans- portation, in: Masters Abstracts International, 2008.
24. P. Gateau, I. Hénaut, L. Barré, J. Argillier, Heavy oil dilution, Oil Gas Sci. Tech(2004) 503e509.
25. P. Luo, C. Yang, Y. Gu, Enhanced solvent dissolution into in-situ upgraded heavy oil under different pressures, Fluid Phase Equilib. 252 (2007) 143e151.
26. R. Urquhart, Heavy oil transportation-present and future, J. Can. Pet. Resources: https://github.com/ol-s-cloud/heavy-oil-rheology-ml (ML Repository on GitHub)