SPE DSEATS Africa Datathon 2025 - Complete Understanding Guide

What This Competition Is About

You're essentially playing detective with oil well data. You have 20 oil wells, and your job is to figure out which of 5 different underground oil reservoirs each well is connected to, plus determine 6 other characteristics about each well's behavior.

Think of it like having 20 different water taps in a building, and you need to figure out which of 5 different water tanks each tap is connected to, plus describe how each tap behaves.

The Big Picture: What You're Solving

Main Goal: Classify 20 oil wells into 5 reservoirs and determine 6 behavioral characteristics for each well.

Why This Matters: Oil companies need to understand which wells connect to which underground oil deposits so they can:

- Optimize production
- Plan maintenance
- Predict future performance
- Allocate resources effectively

Understanding Oil Wells and Reservoirs (Simplified)

What is an Oil Reservoir?

Think of an oil reservoir as an underground "sponge" made of rock that holds oil. It's not a cave or pool - it's more like oil trapped in tiny spaces between rock particles, similar to how a sponge holds water.

What is an Oil Well?

An oil well is like a straw that goes down into the ground to suck oil out of the reservoir sponge. The well connects the surface to the underground oil deposit.

Key Concept: Multiple Wells, Multiple Reservoirs

- You have 20 wells (straws)
- You have 5 reservoirs (underground sponges)

- Each well connects to one reservoir
- Multiple wells can connect to the same reservoir

The 5 Reservoirs You're Working With

Each reservoir has unique characteristics (like different types of sponges):

1. **ACHI**: Medium pressure, moderate oil quality
2. **KEMA**: High pressure, good oil quality
3. **MAKO**: Medium pressure, light oil
4. **DEPU**: Lower pressure, gassy oil
5. **JANI**: Highest pressure, good oil quality

The 7 Things You Need to Classify for Each Well

1. **Reservoir Name** (Which underground sponge?)

- Determine if the well connects to ACHI, KEMA, MAKO, DEPU, or JANI
- **Method**: Compare the well's pressure to each reservoir's pressure (should be within 200 psi)

2. **Reservoir Type** (What kind of sponge?)

- **Saturated**: The oil has as much gas dissolved in it as possible (like a carbonated drink that's fully fizzy)
- **Undersaturated**: The oil could hold more gas (like a flat soda that could be more fizzy)
- **Method**: Compare bubble point pressure to initial reservoir pressure

3. **Well Type** (How does the straw work?)

- **Naturally Flowing (NF)**: Oil comes up on its own due to pressure (like a natural spring)
- **Gas Lifted (GL)**: Gas is injected to help push oil up (like blowing air into a straw to push liquid up)

4. **Production Type** (How consistent is the flow?)

- **Steady**: Oil production is relatively consistent over time
- **Unsteady**: Oil production varies significantly (drops by 50%+ every 3-6 months)

5. **Formation GOR Trend** (How much gas comes with the oil?)

GOR = Gas-Oil Ratio (how much gas comes out with each barrel of oil)

- **Above Solution GOR (aSolGOR)**: More gas than expected (oil is very gassy)

- **At/Below Solution GOR (bSolGOR)**: Normal or less gas than expected

6. **Watercut Trend** (How much water comes with the oil?)

Over time, wells often produce more water mixed with oil:

- **Flat**: Water percentage stays roughly the same
- **Increasing**: Water percentage goes up over time
- **Decreasing**: Water percentage goes down over time
- **Combination**: Mix of different trends

7. **Oil Productivity Index Trend** (How efficient is the well?)

This measures how much oil the well produces per unit of pressure:

- **Flat**: Efficiency stays the same
- **Increasing**: Well gets more efficient over time
- **Decreasing**: Well gets less efficient over time
- **Combination**: Mix of different trends

Understanding the Data You Have

Production Data (Daily Records)

- **Production Date**: When the measurement was taken
- **Onstream Hours**: How many hours the well was producing that day
- **Cumulative Oil/Gas/Water**: Total amounts produced up to that date
- **Choke Size**: Size of the valve controlling flow (bigger = more flow)
- **Bottom Hole Pressure**: Pressure at the bottom of the well
- **Bottom Hole Temperature**: Temperature at the bottom of the well
- **Wellhead Pressure**: Pressure at the surface
- **Wellhead Temperature**: Temperature at the surface
- **Annulus Pressure**: Pressure in the space around the pipe

Reservoir Information

For each of the 5 reservoirs, you know:

- **Initial Pressure**: Starting pressure when first discovered

- **Bubble Point Pressure**: Pressure at which gas starts coming out of oil
- **Current Average Pressure**: Current pressure in the reservoir
- **Solution GOR**: How much gas is dissolved in the oil
- **Oil Formation Volume Factor**: How much oil expands when brought to surface

Your Machine Learning Approach

Step 1: Data Exploration

- Look at the data patterns
- Create graphs showing oil, gas, and water production over time
- Identify trends and anomalies

Step 2: Feature Engineering

Calculate meaningful metrics from raw data:

- **Daily production rates** (not just cumulative)
- **GOR values** (gas production ÷ oil production)
- **Water cut** (water production ÷ total liquid production)
- **Productivity index** (oil rate ÷ pressure drawdown)

Step 3: Classification Logic

Use the rules provided to classify each well:

For Reservoir Assignment:

- Compare each well's maximum bottom hole pressure to each reservoir's current pressure
- Assign to reservoir where pressure difference is ≤ 200 psi

For Other Classifications:

- Calculate trends over time
- Apply the specific rules for each category

Step 4: Model Building

- Use machine learning algorithms (Random Forest, SVM, etc.)
- Train on engineered features
- Validate accuracy

Practical Tips for Success

Data Analysis Strategy

1. **Start with visualizations**: Plot production trends for each well
2. **Calculate key ratios**: GOR, water cut, productivity index
3. **Identify patterns**: Look for steady vs. unsteady production
4. **Apply business rules**: Use the pressure matching for reservoir assignment

Common Pitfalls to Avoid

- Don't create your own classification categories
- Don't modify data outside your Python code
- Don't use AI-assisted code generation
- Don't exceed 10 slides in your presentation

Key Success Factors

1. **Accurate pressure matching** for reservoir assignment
2. **Proper trend analysis** for behavioral classifications
3. **Clear visualizations** in your presentation
4. **Correct total oil calculations** for each reservoir

Deliverables Checklist

Required Files

1. **Google Colab/Jupyter Notebook** (.ipynb) with your ML model
2. **PowerPoint Presentation** (max 10 slides)
3. **Classification Results** (.csv file)

Presentation Structure

1. **Title Slide**: Team name, members, SPE numbers
2. **Outline**: Brief overview of approach
3. **EDA Slides**: Data exploration and visualizations
4. **Methodology**: Your ML workflow
5. **Results**: Classifications and total oil per reservoir
6. **Contributions**: What each team member did

Email Submission

- **To**: speafricadseat@gmail.com
- **Subject**: TeamName_SPE DSEATS Africa 2025 Datathon Submission
- **Deadline**: Monday, July 21, 2025, 11:59 PM WAT

Expected Outcomes

By the end of your analysis, you should have:

- Assigned each of the 20 wells to one of the 5 reservoirs
- Classified each well across 6 behavioral dimensions
- Calculated total oil production from each reservoir
- Created a compelling presentation of your methodology and results

Final Tips for Success

1. **Focus on the pressure matching first** - this is your key to reservoir assignment
2. **Understand the trend definitions** - flat, increasing, decreasing, combination
3. **Validate your classifications** - do they make geological sense?
4. **Tell a clear story** - your presentation should be easy to follow
5. **Double-check your calculations** - especially the total oil per reservoir

Remember: This is fundamentally a pattern recognition problem. You're looking for signatures in the data that indicate which reservoir each well connects to and how each well behaves. The machine learning is just a tool to help you identify and classify these patterns systematically.