SPE DSEATS AFRICA - DATATHON 2025

Guidelines for all Participating Teams

Table of Contents

1.0 General Guidelines

1.1 Overview

The Society of Petroleum Engineer (SPE) Data Science and Engineering Analytics Technical Section (DSEATS) Africa Datathon Challenge presents an exciting opportunity for participants to harness the power of machine learning to address real-world challenges in the oil and gas industry. Organized by SPE DSEATS Africa Region, this challenge aims to leverage historical production data to accurately classify a group of 20 wells based on their observed performance trends.

1.2 Challenge Objective

Participants are expected to work in teams of 3 to 5 persons (max) from at least 2 different organizations and/or schools to develop a machine learning (ML) model that accurately categorizes the 20 wells provided in the dataset by analysing their daily production data trends. The teams should explore innovative approaches to accurately categorize or classify the wells.

1.3 Key Components

- Historical Production Data: Participants will have access to a dataset containing historical production records, including production date, onstream hours, cumulative oil, gas and water production, choke size, bottom hole pressure and temperature, wellhead pressure and temperature and annulus pressure.
- Evaluation Criteria: Submissions will be evaluated based on the accuracy and reliability of the classification models, the creativity and effectiveness of the approaches employed as well as the level of collaboration amongst team members.

1.4 How to Participate

Step 1: Register individually using the link below

https://bit.ly/DSEDATA2

- Interested individuals can register for the challenge by clicking on the registration link above and thereafter shall be provided with the dataset and competition guidelines.

Step 2: Go through the provided datasets, and the submission guidelines

- Please read the general guideline and the submission guideline carefully and be sure to understand and accept the conditions stated in the guidelines. Teams who violate the guidelines will be automatically disqualified from the competition.

Step 3: Build ML model to analyze the datasets provided and make classifications.

- Analyse the datasets provided following the standard machine learning workflow which should include (but not limited to): data preparation, exploratory data analysis, data preprocessing, feature engineering, modelling and classification.

Step 4: Submit your results as a team as per the Submission Guidelines below. Towards the end of the selection process, the top teams being considered for the awards may be invited for an online interview.

1.5 Eligibility

- All members of Society of Petroleum Engineering (SPE) affiliated with any of the SPE Sections in Africa are eligible to participate.
- Past winners of SPE DSEATS Africa Region datathon are eligible to participate however not more than one past winner can join a team.
- Close family and friends of SPE DSEATS Africa board members are not eligible to participate.

2.0 Submission Guidelines

- Participating teams are expected to develop a machine learning model in **Python** programming language to accurately classify the group of 20 wells provided in the dataset. The ML model should be developed in the format of a Google Colab or Jupyter Notebook (.ipynb) file.
- The Google Colab file should be accessible and submitted alongside with the well classification (in csv format) and presentation slides summarizing the work done.
- Submission shall be on a team basis - individual submissions will NOT be accepted. Registered participants are expected to collaborate and form teams of 3 to 5 persons (max) from at least 2 different organizations and/or schools. Each team should adopt a unique but decent, single-phrase team name having not more than 15 letters/characters.
- Teams are expected to save the well classifications in a csv file with the following naming convention: ***TeamName_DSEATS_Africa_2025_Classification.csv***. Adopt similar naming convention for both the PowerPoint slide and Google Colab / Jupyter Notebook file.
- Each team is expected to create a brief PowerPoint presentation of not more than ten (10) slides in total including the Title, Outline, Experimental Data Analysis (EDA), Methodology, Results, Contributions and Thank You slides. A "Thank you" slide is not necessary and can be omitted.
  (Note: Presentation slides that are more than 10 may result in deduction of points which will affect your team's overall score)
- The **Title** slide of your team's presentation should be captioned (e.g *Classification of Oil Wells in the DSEATS Field Using Machine Learning*), followed by your team's name, names of team members and their respective SPE numbers, SPE Sections in Africa that they are affiliated to, School or Company affiliations and date.
- The **EDA** slide(s) should show relevant statistical plots including well production profiles
- The **Methodology** slide(s) should show the ML workflow used to achieve the classification.
- The **Results** slide(s) of your team's presentation should amongst other things show the estimate of the total oil produced (in reservoir barrels) from each of the 5 reservoirs.

- The **Contributions** slide should summarize the work done by each team member.
- The PowerPoint presentation slide deck, the Google Colab file (or Jupyter Notebook file), and the Well Classification file in csv format should be sent to the email address below on or before the submission deadline: **Monday, 21 July 2025 by 11:59PM WAT**.

  Email: ***speafricadseat@gmail.com*** (Note that submissions after the deadline will not be accepted)
- The Subject of the email should read as follows.

  Team's Name_"*SPE DSEATS Africa 2025 Datathon Submission*"
  e.g *Innovisors_SPE DSEATS Africa 2025 Datathon Submission*

  The body of the email should be in simple english as shown below.

  Dear Reviewer,
  On behalf of team (name), find attached to this mail, presentation slide deck, the Google Colab file containing our solution, and the well classification .csv file.

  Note:
  a. Please, do not forget to attach the files (Presentation Slides, the Google Colab File, and the Well Classification .csv file)
  b. Only one member of the team should make the submission on behalf of the team and copy the other team members. Multiple submissions by different team members will attract penalties and possible disqualification.
  c. Only one final submission will be accepted, pls indicate which is your team's final one if you have made more than one submission. Not more than 2 submissions in all should be made over the competition period. Teams making more than 2 submissions will be disqualified.

3.0 Dataset, Reservoir Information & Expectations

The dataset provided is a DSEATS conceptual field with historical production information such as: onstream hours, cumulative oil, gas and water production, choke size, bottom hole pressure and temperature, wellhead pressure and temperature and annulus pressure (all in Field Units). Registered participants are provided with:

1) **Wells dataset:** There are twenty (20) oil wells in the dataset attached with the email named: ***spe_africa_dseats_datathon_2025_wells_dataset.csv*** file
2) **Reservoir information:** There are five(5) black oil reservoirs with relevant information as shown in the table below and also attached with the email as ***reservoir_info.csv*** file

Reservoir Name	Initial Reservoir Pressure (PSI)	Bubble Point Pressure (PSI)	Current Average Reservoir Pressure (PSI)	Solution Gas Oil Ratio (SCF/BBL)	Oil Formation Volume Factor (RB/STB)
ACHI	3500	3300	2700	800	1.20
KEMA	4200	4000	3900	600	1.45
MAKO	3500	3500	3000	500	1.15
DEPU	2800	2800	2400	1200	1.37
JANI	4500	4300	4200	1000	1.30

3) **Well Classification Parameters**: There are seven (7) classification parameters as shown in the table below and also attached with the email as ***classification_parameters.csv*** file

Reservoir Name	Reservoir Type	Well Type	Production Type	Formation GOR Trend	Watercut Trend	Oil Productivity Index Trend
ACHI	Saturated	NF	Steady	aSolGOR	Flat	Flat
KEMA	Undersat	GL	Unsteady	bSolGOR	Incr	Incr
MAKO				Combo	Decr	Decr
DEPU					Combo	Combo
JANI						

NF: Naturally Flowing
GL: Gas Lifted
Undersat: Undersaturated reservoir
aSolGOR: Above Solution GOR
bSolGOR: At or below solution gas oil ratio (GOR)

Flat: Flat or fairly horizontal line trend
Incr: Increasing trend
Decr: Decreasing trend
Combo: Combination of two or more trends

Instructions / Expectations:

Using the dataset and the reservoir information provided, the teams are expected to develop Python code to:

1) Identify the reservoir each well is producing from assuming not more than 200 psi differential pressure at the maximum bottom hole pressure.
2) Determine for each reservoir whether saturated or undersaturated using the initial condition
3) Identify whether the well is naturally flowing or gas lifted
4) Determine if production from a well is steady or unsteady. Consider unsteady wells as those with more than 50% drop in oil production at least once every 3 to 6 months
5) Determine from the formation GOR trend whether the well is flowing at/below or above the solution GOR of the selected reservoir or a combination of both trends
6) Determine from the watercut trend whether it is Flat, Increasing, Decreasing or a Combination of 2 or more trends
7) Determine from the oil productivity index trend whether it is Flat, Increasing, Decreasing or a Combination of 2 or more trends. You may use average reservoir pressure for the oil productivity index computation
8) Teams should show the relevant production profiles / trend plots as part of the EDA in their Python code
9) Use ONLY the classification categories provided - DO NOT create your own
10) Import the dataset and .csv files as provided, do NOT manipulate outside your Python code
11) The classification results should be saved in a .csv file and submitted alongside the PowerPoint slide deck and Google Colab file. A sample of the expected results file called: ***Innovisors_DSEATS_Africa_2025_Classification.csv*** has been attached in the email for reference purposes only and as such the classification of the first well is intentionally not correct. Kindly comply with the column order and provide your team's classifications for all the wells from 1 to 20.
12) Calculate the total reservoir barrels of oil produced from each reservoir and include the results in the **Results** slide(s) of their PowerPoint presentation slide deck.

4.0 Evaluation Criteria

Judges will assess the effectiveness of the machine learning techniques employed and the potential impact of the predictive models on production optimization in the oil and gas industry. The following are the criteria that would be used.

- Effectiveness of the ML Techniques: Judges will evaluate the efficiency of the machine learning techniques / algorithms used.

- Classification Accuracy: The classification model's accuracy and performance would be evaluated with key performance indicators such as Accuracy, Precision, Recall, F1-Score etc
- Total Production Estimation: Teams are expected to provide the total reservoir barrels of oil produced from each reservoir and include the results in the Results slide(s) of their PowerPoint presentation slide deck.
- Innovation and Creativity: Additional points shall be given to teams who adopt innovative and creative approaches in applying machine learning to solve this challenge. However, please note that the use of Artificial Intelligence (AI)-assisted techniques or Generative AI is highly discouraged and will result in disqualification of any team involved. All submissions would first be curated using our AI detection tool.

5.0 Intellectual Property

The teams retain ownership of their submissions; however, by participating in the competition, participants grant the organizers the right to showcase and promote their work within the SPE community

6.0 Rewards & Benefits

The best three (3) teams shall win cash prizes amongst other benefits.

- Cash Prize: The competition promises to reward the best outstanding teams with some take home cash prize.
- Awards: The top 3 teams shall be recognised at the Awards Event holding during the SPE NAICE at Eko Hotel & Suites in Lagos, Nigeria from August 4th to 6th 2025.
- Recognition: Outstanding submissions will be recognized and showcased within the SPE community, offering participants visibility and professional recognition in SPE Africa, and SPE International.
- Professional Development: Participating teams will have the opportunity to enhance their machine learning skills, gain practical experience in data analysis, and network with industry experts.