DATATHON 2024

GUIDELINES FOR ALL PARTICIPANTS

(GENERAL GUIDELINES & SUBMISSION GUIDELINES)

Table of Contents

1.0 General Guidelines

1.1 Overview

The SPE Data Science and Engineering Analytics Technical Section (DSEAT) 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 the Society of Petroleum Engineers (SPE) Data Science and Engineering Analytics Technical Section (DSEAT) Africa, this challenge aims to leverage historical production data to predict future production performance.

1.2 Challenge Objective

Participants are to develop a machine learning model to predict oil, gas, and water production based on historical records.

By analysing past production data and associated parameters, participants should apply the standard machine learning (ML) workflow to accurately forecast production profile of the oil, gas, and water.

1.3 Key Components

- Historical Production Data: Participants will have access to a dataset containing historical production records, including production rates, choke sizes, and tubing head pressure measurements et cetera.
- Evaluation Criteria: Submissions will be evaluated based on the accuracy and reliability of their predictions, as well as the creativity and effectiveness of the machine learning approaches employed.

1.4 How to Participate

Step 1: Register for the completion using this link: https://bit.ly/3KrKFVF

- Interested SPE members (Students and Young Professionals) can register for the challenge and will subsequently be provided with the datasets and competition guidelines. By clicking on the registration link above

Step 2: Read through this general and submission guidelines document.

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

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

- Analyse the training and validation datasets provide following the standard machine learning workflow which should include (but not limited to): data preparation, exploratory data analysis, feature engineering, data splitting, modelling and prediction.

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

1.5 Eligibility

- Students and Young Professionals who are current members of Society of Petroleum Engineers (SPE) affiliated to any of the SPE Sections in Africa are eligible to participate.

Note: *Past African Regional Winners of SPE Data Science and Engineering Analytics, and SPE Africa DSEAT Committee members are not eligible to participate*

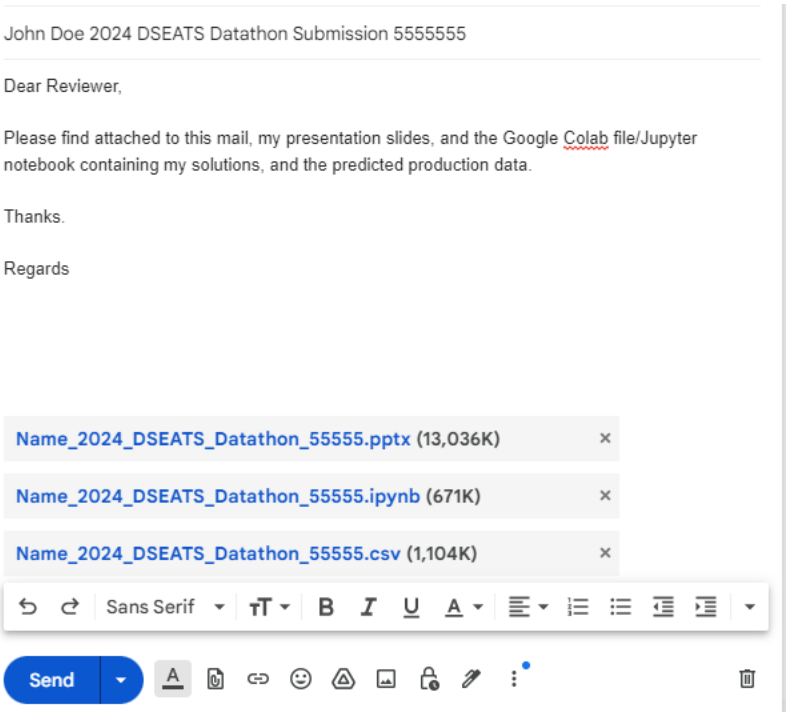
2.0 Submission Guidelines

1. Participants are expected to develop a machine learning model using **Python** programming language, to predict the oil, gas and water production. The ML model should be developed in the format of a Google Colab or Jupyter Notebook (.ipynb) file.
2. The Google Colab file should be accessible and submitted alongside with the predicted production data and PowerPoint slides showing the work done.
3. Participants are also expected to save the predicted production information in a .csv file as: *Firstname_Surname_2024_DSEATS_Datathon_SPE Number.csv.* Adopt similar naming convention for both the PowerPoint and Google Colab / Jupyter Notebook file.
4. Participants are also expected to create a brief PowerPoint presentation of not more than ten (10) slides in total including the Title, Outline, Results and Thank You slides. A "Thank you" slide can be omitted. The presentation slides should show the ML workflow/steps, and procedures used to achieve the predictions.
   (Note if presentation slides are more than 10 this may result in deduction of points which will affect your overall score)
5. The Title page of your presentation should show: the title (e.g Prediction of Oil, Gas & Water Production in the DSEATS Field Using Machine Learning), your first name and surname, SPE number, SPE Section in Africa you are affiliated to, Category (Student or Young Professional), Course of Study/Technical Area, School or Company affiliation and date.
6. The PowerPoint presentation slides, The Google Colab (or Jupyter Notebook) file, and the Predicted Production Data file in .csv format should be sent to the email address below on or before the submission deadline: **Saturday, 13 July 2024 by 11:59PM WAT**.
   Email: ***speafricadseat@gmail.com***
   (Note that submissions after the deadline will not be accepted)

7. The Subject of the email should be used as shown below.
   Firstname Surname “2024 DSEATS Datathon Submission” SPE Number
   e.g John Doe 2024 DSEATS Datathon Submission 5555555

   The body of the email should have a simple text as shown below.

   Dear Reviewer,
   Please find attached to this mail, my presentation slides, and the Google Colab file containing my solution, and the predicted production data.



Note:

- **a.** Please, do not forget to attach the files (Presentation Slides, The Google Colab File, and The Predicted Production Data). Incomplete/partial submissions will not be accepted.
- **b.** If you are sending a link to a Google Colab file, please ensure it is accessible by enabling the permission, you can test this by sending it to a colleague see if they can open it, before inserting it to the mail.
- **c.** Only one final submission will be accepted, pls indicate which is your final one if you have made more than one submission. Not more than 2 submissions in all should be made over the competition period. Candidates making more than 2 submissions will be disqualified.
- **d.** Kindly use the same email address used for registration to submit your work.
- **e.** Any submissions made after the submission deadline would not be accepted

3.0 Datasets Provided

The production data provided is a DSEATS conceptual oil field with historical production information such as Well Code, Field Name, Downhole Pressure, Tubing Head Pressure, Choke Size et cetera. (Units are in Field Units).

Participants are given two data files:

1) Training dataset
2) Validation dataset

Using the features in the training dataset, participants are expected to develop their best machine learning model and then use this to predict or forecast the oil, gas, and water production in the validation dataset.

4.0 Evaluation Criteria

The judges will assess the effectiveness of the machine learning techniques employed and the accuracy of the oil, gas and water predictions. The following are the criteria that would be used.

- Effectiveness of the ML Techniques: Judges will evaluate the effectiveness of the machine learning techniques used.
- Prediction Accuracy: The predictive model's accuracy and performance would be evaluated with key performance indicators such as **RMSE, R2, MAE** et cetera.
- Total Production Estimation: Participants are expected to provide estimates of the total oil, gas and water production (over the history and forecast periods). These estimates should be provided in the Results slide in the PowerPoint slide pack.
- Innovation, and Creativity: Additional scores would be given to participants who adopt innovative and creative approach in applying machine learning to solve this challenge. However, please note that the use of Artificial Intelligence (AI)-assisted techniques such as Chat-GPT, Gemini, Mistral-AI et cetera is highly discouraged and will result in disqualification of the participant.

5.0 Intellectual Property

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

6.0 Rewards & Benefits

The top (3) participants in each category (Students Category and Young Professional Category) would win cash prizes, and other awards.

- Cash Prize: The competition promises to reward the best outstanding participants with some take home cash prize.
- Awards: Participants would be recognised and awarded at the 2024 SPE NAICE Conference holding in Nigeria, at Eko Hotel & Suites Lagos from 5 to 8 August 2024.
- 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: Participants will have the opportunity to enhance their machine learning skills, gain practical experience in data analytics and network with industry experts.