Assessment

An oil palm plantation estate has engaged you as a Data Scientist to develop a model that predicts the production of their oil palm estate. After discussing with the stakeholders, you are required to develop a model that **predicts monthly oil palm production** for **January, February and March 2024**. The following key information were shared by the Subject Matter Expert related to oil palm production:

- **Production** data have yearly seasonality. The duration of oil palm fruit production takes up to 36 months.
- Production is associated with its **age** (growth), with no production from age 0 – 3, low production slowly increasing from age 4 – 8, optimal production from age 9 – 15, plateauing production from age 16 – 20, and drop in production from age 21 onwards.
- Production is affected by how many **palms is planted in an area**. Generally a bigger area would have more palm stand.
- **Fertiliser** is a key important input for oil palm to produce fruit. Fertiliser is applied throughout the year with only some month having fertiliser applied. Study have shown fertiliser applied up to past 24 months ago can affect the production.
- **Rainfall** have major effects on production, in deficit or excessive amount. Monthly rainfall exceeding 300mm or less than 100mm for consecutive 3 months will have severe effect on the production. Study have shown that past rainfall up to 24 months ago can affect the production.

Based on the information gathered, the following **datasets collected in monthly basis** from January 2013 to December 2023 was provided:

- Date: The date when the data is collected
- Production: The historical production (in metric tonne)
- Area: The size of the land (in hectarage) planted with oil palm
- Age: The age of the palm (in year)
- Palm Stand: The number of palms planted in the area
- Fertiliser: The amount of fertiliser applied (in kg/palm)
- Rainfall: The amount of rainfall (in mm)
- Rainday: The amount of rain days

With the information above, **you are required to**:

1. Carry out necessary data wrangling on the provided dataset
2. Develop model(s) using suitable machine learning algorithm to predict monthly production for January, February and March 2024
3. Use suitable metrics to measure the model performance
4. Provide explanation of the predicted value

The following are to be provided (in word or visualisation):

1. Explanation on your data wrangling procedure
2. Explanation on your selected metrics and model (with support from your selected metrics)
3. Explanation on the features affecting production
4. Predicted values and explanation for January, February and March 2024

Please ensure to provide an executable file or a script file so that the codes could be tested on another windows-based platform.