Data Challenge DSW 2023

Senior Professional Category

Problem Statement

A fertilizer company wants to maintain its product catalogs. The information from product catalogs are taken from various sources, such as point of sale (POS) transactions. Sometimes, the company's partners put free-text product names to the POS transaction system. The fertilizer company wants to **map the free-text input product names with its product catalog's stock keeping unit (SKU)**. Given the attached dataset, you are challenged to build a solution that can map the product name from POS transactions with the product SKU from product catalogs. There are several solutions that you can build (you may want to select one or more, or even propose other solutions based on the given dataset), such as:

1. Map the **product names** from POS transactions to **product SKU**. You can use methods, such as **similarity,** or any methods to map the product names. If the product name from POS transactions does not exist in the product catalog or have low similarity with the maintained product SKU, you can propose a new product SKU in the product catalog. It is **optional** to map the free-text input product names to the product catalog's brand, type, and formula. It is also **optional** to build an analysis report of the product catalog in the form of a dashboard or analysis report. Participants are also encouraged to use external dataset to enrich the analysis.
2. Build a **model** that can predict the product SKU in the product catalog by the product name of the POS transaction. If the product name from the POS transaction does not exist or have low similarity, the **model** can propose a new product SKU in the product catalog.

Dataset Description

There are two datasets provided: **Product Catalog** and **Product Name from POS transactions**. There are several columns provided in the dataset:

- **Product Catalog**
    1. **Product SKU**
    2. **Brand**

3. **Type**
4. **Formula**

- **Product Name from PoS Transactions**
    1. **Product Name**

Scoring

In the preliminary round, your result will be scored based on these qualities:

- Originality
- Method suitability to the problem provided
- Impact

In the final round, in addition to the aforementioned scoring criteria, the presentation will also be evaluated.

Solution Submission

- Alongside with the solution, participants are required to submit a PowerPoint presentation file in the format of:
    1. **Title**
       A short title describing the issues raised and the solutions provided.
    2. **Team Introduction**
       Team name and the educational and professional background of the team members.
    3. **Introduction**
       The background of the topics and issues raised, including the urgency and importance of the issues.
    4. **Data & Methodology**
       An explanation of the datasets, steps, and tools used to analyze and solve the problems.
    5. **Results**
       A set of essential findings from the analysis, including metrics (if any).
    6. **Conclusion**
       Analysis summary and solutions proposed for the raised issues.
- Participants must follow the submission format as follows: **“DataChallenge_(Team Name).zip”**. The .zip file consists of the **PowerPoint**

presentation file and the **solution** to the given topics and datasets. For example, if the solution is in the form of a public Tableau dashboard, participants may mention and attach the link to the public Tableau dashboard in the PowerPoint presentation file. Participants may also attach the solution in the zipped file if the solution cannot be accessed via a public website.

- Final submission must be uploaded to this submission link: https://bit.ly/SolutionSubmissionDSW.
- Participants are allowed to change or update their solutions until the registration closes on **12 November 2023 at 11:59 PM (23:59) GMT+7 (Jakarta Time)**.