SMARTBatch Project Documentation

I. Project Introduction:

L'Oréal is dedicated to upholding high standards of efficiency and quality in its production processes. However, the company's factories have been experiencing significant inefficiencies due to frequent adjustments required during the production phase of certain formulas. These adjustments have led to increased production time, higher resource consumption, and inconsistencies in product quality. To address these challenges, the SMARTBatch project was launched with the goal of developing a machine learning system that recommends new formula references, thereby minimizing the need for batch adjustments. This initiative aims to enhance production stability and consistency across all L'Oréal factories. Currently, the project is being tested at the **Lassigny factory**, focusing on lipstick formulas, with plans for future expansion to all L'Oréal brands and product lines.

II. Problem Description:

The following bar graphs illustrate the production issues faced by the Lassigny factory with various formulas. The first graph shows the relative percentages of raw materials in the lipstick products, highlighting variations in distribution between different batches. The second graph depicts the number of adjustments made per batch and the percentage of batches that required adjustments for each formula. This data clearly illustrates the inconsistencies and the need for adjustments across different formulas, underscoring the inefficiencies in the current production process.

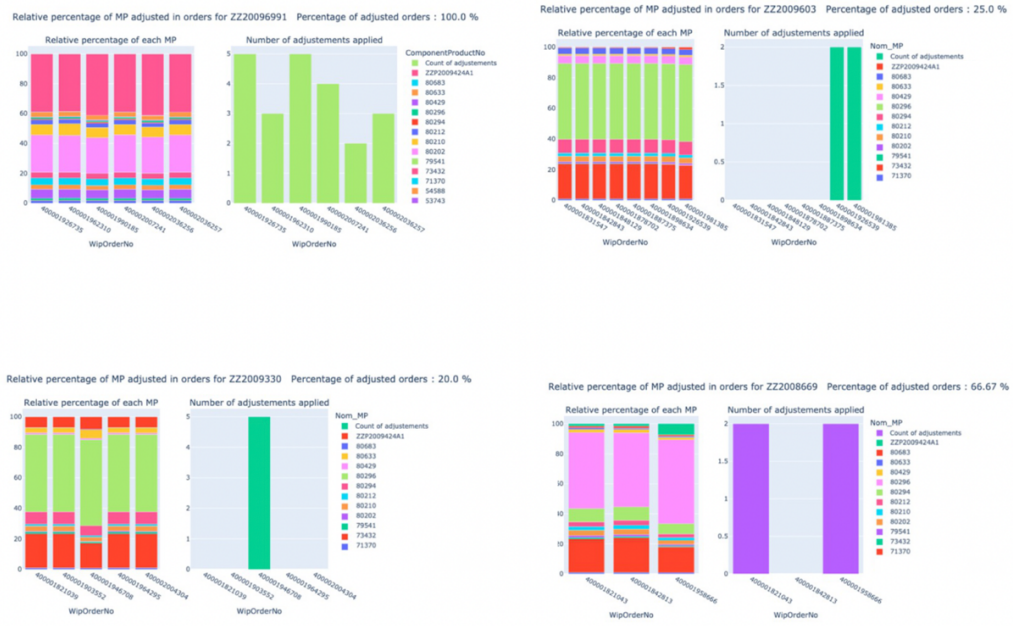

Figure 1 Bar graph figures describing some effected formulas

- Key Data Fields provided by the factories:
  - WipOrderNo: Batch ID
  - N formulas: Formula ID
  - ComponentProductNo: Raw Material Number
  - Intro Pesée en Gr: Initial Weight of Raw Material
  - Ajust Pesée en Gr: Adjusted Weight of Raw Material
  - Skid: Container Used
  - Date Intro: Initialization Date
- Data Limitations:
  - Limited number of batches per formula, affecting the performance of machine learning models.
  - Insufficient data granularity regarding raw material suppliers and purchase dates.

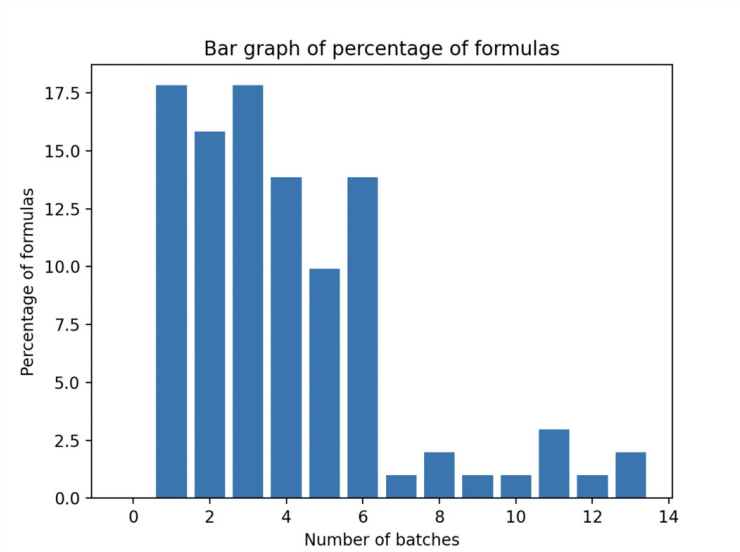

Figure 2 Bar graph presenting percentages of formulas for each number of batches

III. Project Process:

Comprehensive data analysis was conducted to identify the inefficiencies caused by frequent formula adjustments. Initial efforts involved benchmarking various machine learning models to predict new formula references to streamline production. However, these models showed low performance, primarily due to insufficient data volume and lack of data granularity.

To overcome these challenges, a shift was made to simpler, rule-based techniques developed in collaboration with business stakeholders. This approach effectively addressed the problem and improved production processes. The strategy to enhance production efficiency and consistency involves a two-stage process:

A. Classification Stage:

This initial phase focuses on identifying formulas that have recently achieved stability. By classifying these stable formulas, we can determine which ones exhibit consistent performance and are less likely to require frequent adjustments.

1. **Filtering Batches Based on Stability Criteria:** Apply filters to batches of formulas using predefined constraints, known as Formulas Stability Criteria. These criteria include:

    - Number of Recent Batches Considered: Specify the quantity of recent batches to analyze.
    - Minimum Accepted Batches: Define the minimum number of batches required for reliable stability assessment.
    - Threshold Number of Adjustments Accepted: Set a limit on the number of adjustments permitted within the analyzed batches.

2. **Assesing Variance for Stability Classification:** Evaluate the formulas to determine if they meet the threshold for the coefficient of variation parameter. This step ensures that only those formulas with stable batch performance, as indicated by low variance, are classified as stable.

B. Gridsearch Filtering Stage:

The second phase involves optimizing parameters to refine the filtering process for batches derived from formulas identified as unstable. The objective is to identify the optimal combination of filtering parameters that minimizes the mean absolute error (MAE) in predicting the composition of the latest batch. This stage employs grid search techniques to systematically evaluate and select the most effective parameters for improving batch stability and reducing the need for adjustments.

1. **Parameters to optimize:**

    - *Number of Outlier Batches to Remove*: Filtering outliers to improve model accuracy.
    - *Minimum Accepted Batches After Removing Outliers*: Ensuring sufficient data for reliable predictions.
    - *Maximum Number of Adjustments Allowed*: Limiting the number of adjustments to enhance stability.
    - *Initialization Date Limit*: Considering only batches after a specific date to ensure relevance.
    - *Weighted Calculations*: Whether to apply weights to different batches (true or false).
    - *Initial Weight and Increase Rate*: Setting initial weights and their rate of increase for weighted calculations.

2. **Evaluation Metrics:**

    - Mean Absolute Error (MAE): The primary metric for assessing the accuracy of batch predictions.
    - Ground Truth: Utilize the latest batch as the reference to measure prediction accuracy.

3. **Results Interpretation:**
   The results of the optimization are assessed by examining the average relative percentages of each raw material in the selected formulas, based on the batches filtered through the grid search process. This analysis provides insights into the stability and composition of the formulas, thereby validating the effectiveness of the parameter optimization.

IV. Project Output:

The project has resulted in the creation of a comprehensive Excel file that includes:

- A detailed list of both new and existing formula references, highlighting significant differences in raw materials and providing recommendations based on the new references, facilitating comparison and integration.

- Information on detailed specifications, including the parameters used for filtering, performance metrics, and prioritization guidelines to support production decision-making