**Particle Formation Use Case**

**Description:**

One of Covestro’s core product classes are dispersions used for coatings and specialty films. To sell a high grade dispersion, the product must meet ridge quality criteria. Depending on the application, only a certain number and size of particles are acceptable. This quality measure is used to ensure customer satisfaction and, in principle, smaller and fewer particles leads to a higher sales price. The quality score is also needed to apply an automated root-cause analysis for explaining the causal relationship between input parameters and the quality of the dispersion. Therefore, a stable and reliable quality measure is key. Currently, the process of evaluating the film quality is as follows:

1. Dispersion is applied to a clean surface.
2. An image is taken and processed.
3. Multiple quality managers score the image based on their experience.
4. The ratings are combined to determine an overall score.

**Pain Point:**

Due to poor image quality and the high subjectivity of the scoring process, quality measurements of the dispersions are inconsistent and weak with regards to reproducibility. Although quality managers at Covestro are highly skilled, their scoring results have been proven unstable. A quality manager that rates the same image at two different points in time sometimes assigns a quite different quality score.

A stable process is needed to provide reliable results that can be used to draw conclusions that can help to increase the product quality in the long-term.

**Data Provided:**

A data set containing dispersion images that have been rated by multiple quality managers is provided. On the right side of each image, there is a section of classified information that has been purposefully hidden by a black bar. This information is not relevant to your analysis. The images contain the mentioned particles, which vary in number and size.

**Goal and Evaluation Metric:**

The goal of this challenge is to develop an automated and consistent method to rate the quality of a dispersion. Performance of the model will be evaluated using R2.

**Submission File:**

The submission file should contain the predicted quality ratings for each entry in the to\_predict.csv file. The submission should be a CSV file entitled “team**#**submission.csv”, where **#** is your assigned team number, and should follow the template format provided for the chosen use case. The file should be sent to the following email address: [Chemalytix@covestro.com](mailto:Chemalytix@covestro.com).

**Data Details**

You are provided two folders, “train” and “test” containing the images for training and testing, respectively. For reference, the following two images can be used as a baseline for what is considered a high quality and low quality dispersion.

|  |  |
| --- | --- |
| **High Rating Image:**  **C:\Users\ggeyt\Downloads\1452a179d2d198c20b1c84a15881f47c848ffb63.bmp** | **Low Rating Image:**  **C:\Users\ggeyt\Downloads\fe16738489559f7003793531eb9a67797b50ae6d.bmp** |

Details for the remaining provided files are as follows:

**train\_scores.csv**

|  |  |
| --- | --- |
| Variable | Description |
| image | image ID (corresponds to file name) |
| exp<i> | rating of expert quality manager <i> |
| total\_rating | rating of the dispersion derived from expert rating (target variable) |

**to\_predict.csv:**

Data points to be predicted.

|  |  |
| --- | --- |
| Variable | Description |
| image | ID of image to use for prediction |

**team#submission.csv**

Sample submission file representing the expected format of the submission file.

|  |  |
| --- | --- |
| Variable | Description |
| image | image ID (corresponds to file name) |
| total\_rating | predicted rating of the dispersion (target variable) |