SUPPLEMENTARY MATERIAL

Machine Learning Approach to Predict Physical Properties of Polypropylene Composites: Application of MLR, DNN, and Random Forest to Industrial Data

Chonghyo Joo [1,2,†], Hyundo Park [1,2,†], Hyukwon Kwon [1,2], Jongkoo Lim [3], Eunchul Shin [3], Hyungtae Cho [1] and Junghwan Kim [1,*]

[1] *Green Materials and Processes R&D Group, Korea Institute of Industrial Technology, 55, Jongga-ro, Jung-gu, 44413 Ulsan, Korea*

[2] *Department of Chemical and Biomolecular Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, 03722 Seoul, Korea*

[3] *Research & Development Center, GS Caltex Corporation, 359, Expo-ro, Yuseon-gu, 34122 Daejeon, Korea*

[†] **Co-first authors: These authors contributed equally to this work.**

[*]**Corresponding author**

Junghwan Kim

E-mail: kjh31@kitech.re.kr

Table S1. Pseudo code of categorization

Algorithm: Categorization

Input: $\mathcal{M}$: n × d recipe dataset, n: Number of recipes, d: Number of materials.

Output: C: Category set.

1: Binning process

2: **for each** $u \in \mathcal{M}$ **do**

3: **if** $u > 0$ // If u is bigger than zero, replace the value of u with "T."

4: replace u value with "T"

5: **if** $u = 0$ // If u is zero, replace the value of u with "F."

6: replace u value with "F"

7: **end**

8: $\mathcal{M} \rightarrow \mathcal{M}'$ // Recipe dataset replaced with "T" and "F" after binning process.

9: Categorization process

10: **for each** $u \in \mathcal{M}'$ **do**

11: arrange u from u (n, 1) to u (n, d) // Make code a consisting of "T" and "F."

12: give the arranged code a to data point (n,) // Give code a to each recipe.

13: **for each** code $a \in \mathcal{M}'$ **do**

14: **for each** code $b \in \mathcal{M}' \setminus u$ **do**

15: **if** $a = b$ // If a and b are same, add the recipes to the same category

16: add (a, b) pair to the same category in the category set C

16: **Return** C