**Classification of Rakuten e-commerce products**

**Relevant training course:** Data Scientist

**Level of difficulty:** 07/10

**Description of the project:**

*Cataloging products according to different data (texts and images) is important for e-commerce since it allows for various applications such as product recommendation and personalized research. It is then a question of predicting the type code of the products knowing textual data (designation and description of the products) as well as image data (image of the product).*

**Resources to refer to:**

* **Data:** 
  + This project is part of the Rakuten France Multimodal Product Data Classification challenge, the data and their description are available at: <https://challengedata.ens.fr/challenges/35>
  + Text data: ~60 mb
  + Image data: ~2.2 gb
  + 99k data with over 1000 classes.

**Validation conditions for the project:**

* an exploration, data visualization and data pre-processing **report** ;
* a modeling **report** ;
* a final **report** and associated **GitHub**.