



MLE Assignment

This document contains privileged information as property of GfK SE. Please note that unauthorized copying, disclosure or distribution of the material in this document is not permitted.



Python - ML (45 minutes)

Dataset C: [testset_C.csv](#)

Dataset Description:

Dataset C is supposed to contain records with article texts and the belonging product group. The following information is known about the columns:

Column	Info
id	A unique record identifier
product group	Product category
main_text	a describing text about the article
add_text	an additional describing text about the article
manufacturer	the manufacturer belonging to the article

Tasks:

1. Create a machine learning model in order to predict the product category based on appropriate columns. Use scikit-learn with one or more machine learning algorithms.
2. Present the result in a vivid way and explain your model from a statistical PoV.
3. Create a microservice with a REST-API which obtains an article text and returns the product category.
4. Convert all the code written in above sub tasks (1,2 & 3) to production-quality code (Hint: README, test cases etc.)
5. Commit all code & results to Git. Share the git repository with the GfK MLE Team members mentioned in the email.