Data Scientist Case Study

Problem:

In the MyHammer platform we have one challenge. Often consumers submit the service request in a wrong serviceId category, mostly they get tempted to choose easy options such as "Sonstiges" (Others).

For example, an actual painting job gets submitted to "Others" serviceId category.

We have discovered the following issues with the current setup:

- jobs are in a wrong category
- tradesmen cannot find the proper job because of the above scenario.

This has a direct impact in revenue because the tradesmen cannot find a relevant job while searching.

Your task is to build an efficient classifier which can classify the wrong serviceId to a correct one.

The goal is to help consumers to pick the right category by providing a good suggestion of a correct serviceId category based on the service request details.

Please use the attached dataset, the serviceId is your target variable which you should focus on.

Please provide us with a Jupyter Notebook file with your results.