Arquitectura de productos de datos (Datos de gran escala) -HW04

Thomas M. Rudolf

1. Review the following blog on proofs of concept.

https://aws.amazon.com/blogs/enterprise-strategy/generative-ai-getting-proofs-of-concept-to-production/

Aim:

Understand what a POC is. POC stands for **P**roof-**o**f-**C**oncept

What are the objectives of a POC?

Answer: POC is an important tool to evaluate the possible impact and risks (financial or image) of a product. Among other, the costs are analysed, risks can be reduced and the requirements and specifications to build the product can be explored.

Why should a POC be designed in order to put it into production?

Answer: POC is an important tool to evaluate the feasibility, risks and validate the a business case. A POC will not vanish the risks that exist but the evaluation allows to mange risks and most likely to diminish negative consequences.

What should be the results of a successful POC?

Answer: The results of a successful POC should be a good estimation of cost, possible risks and how they can be diminished. An outcome should also be if the product fits into the business work frame and what is the value that value creates this product to the company. Based on this result, a decision can be made if the development continues or not.

How is the POC related to the prototype of the house or sales problem that you carried out in task 02?

Answer: The implemented prototype can give information about the feasibility. Certain risk about false estimation can be obtained considering errors, mean values and especially

standard deviation can be used to analyse the quality. But there is no compatibility check done regarding the business framework, e.g..

What results were you able to obtain from building the prototype on the notebook?

Answer:

The following results were obtained yet:

Mean Absolute Error: 17821.348

Mean Absolute Percentage Error: 0.1

Standard deviation is missing and needs to be analysed.

What findings are going to influence your decisions to turn this into a data product that is usable?

Answer: Currently I am not taking into account the categorical values. I would prefer to include these parameters and calculate the standard deviation as well. Once included this and a positive result, I would start to define a data product.