

d.Data Analytics in Consulting-Target Reply

Synthetic Data

Something we don't know in the past.

We have different way to synthesize elements.

Synthetic data are data that we can create in great quantity and are indistinguishable from real data. A synthetic dataset don't contains reference to the real world, cannot take more value from the data than the one thought in advance. Useful for testing/simulation as it can help to speed up the testing phase. Synthetic data is a better solution for enriching dataset maintaining the underlying characteristics of the dataset.

Helps to explore new realities as it can allow the exploration of new unexplored situations.

How it works

We use a Generative Adversarial Networks: we use a Generator and a Discriminator. The Generator create data to fake the Discriminator and the Discriminator identify the false data given by the Generator. Another way is using Variational Autoencoder: encoding the original data and obtain a model to describe the data and from it you derive a generalized model for the data.

Use cases

- Multipurpose Synthetic data framework
 - Data Monetization: enable data sharing to have a monetary return
 - Fraud Detection: identify fake ID documents
 - Data Masking for Development: use realistic data in products development

Challenges

- Representativeness: ensuring that synthetic data accurately represent the statistical properties and patterns of the original data can be challenging. The generated data must sufficiently capture the complexity and diversity of the real world scenario.
- Validations: validating the effectiveness of synthetic data is crucial. It involves assessing whether models trained on synthetic data generalize well on real-world data, ensuring that the synthetic data truly reflects the underlying patterns.
- Bias and Fairness:
- Transparency:

Explainable AI(XAI)

You have to avoid to be in a state where you can't understand what is going on with the black box you created/you are working with. You have to explain the behaviour of the model really well and in a simple way.

Explainable AI is a field that aim to understand ML algorithm and tries to explain them in a human language.

They can help to understand the complexity of black box model and algorithm. You have to create trust between the parts and you can do it providing useful example of what s going to happens.

Post-hoc approaches are machine learning approaches of machine learning and are used to understand what is happening. They can be applied only to specific ML algorithms. Post-hoc Global model explain all the units in the dataset. Post-hoc Local gives limited explanation to specific kind of dataset. Post-hoc Model specific supports explainability constraints with respect to learning algorithm and internal structure of the model. Post-hoc Model Agnostic applies pair wise-analysis of model inputs and prediction.

These algorithms are useful for explain ti justify, explain to control, explain to discover and explain to improve.