INSIGHTS · INTELIGÊNCIA ARTIFICIAL

Dados sintéticos no Microsoft Fabric: gerar, validar e integrar sem sair do ecossistema

Desliza →

bConcepts
Consulting Services

www.bconcepts.pt

INTELIGÊNCIA ARTIFICIAL

01

Porquê usar dados sintéticos?

Escassez, privacidade e classes raras travam modelos. Sintéticos permitem replicar padrões e acelerar desenvolvimento sem expor dados sensíveis.

INTELIGÊNCIA ARTIFICIAL

02

Tipos e técnicas

Tabular (CTGAN, SDV), séries temporais (TimeGAN, LSTM), texto (LLMs afinados) e imagens (GANs, augmentations) — cada tipo exige técnicas distintas.

INTELIGÊNCIA ARTIFICIAL

03

Fluxo prático em 5 passos

Preparar amostra no OneLake, gerar no Notebook PySpark, validar estatisticamente, avaliar utilidade e operacionalizar com tracking de artefactos.

INTELIGÊNCIA ARTIFICIAL

04

Testes essenciais

KS e Chi-square para marginais, comparação de matrizes de correlação, PCA e verificação de autocorrelações para séries temporais; documentar thresholds.

INTELIGÊNCIA ARTIFICIAL

05

Métricas de utilidade

Comparar treino com real, real+synthetic e apenas synthetic. Ex.: adicionar 150k sintéticos pode aumentar recall de 62% para 72% nalguns casos.

INTELIGÊNCIA ARTIFICIAL

06

Riscos e salvaguardas

Evitar réplicas (distância mínima), aplicar differential privacy quando necessário, auditar vieses por subgrupos e definir regras de pass/fail.

INTELIGÊNCIA ARTIFICIAL

07

Automação e reproducibilidade

Registar seeds, versões de gerador e métricas em MLflow ou repositório do Fabric; automatizar validações antes de promover para produção.

CONTINUA NO BLOG

Como gerar, validar e integrar dados sintéticos no Microsoft Fabric para melhorar modelos, proteger privacidade e reduzir custos operacionais?

www.bconcepts.pt/insights

bConcepts
Consulting Services

Artigo completo → link nos comentários