FinTech Accelerator Proof of Concept

Privitar - Data Desensitisation Tool

Executive Summary

The Bank of England has set up a FinTech Accelerator to work with innovative firms and new technologies. Data desensitisation is one of the areas that the Accelerator is examining. This will allow Bank researchers wide access to data collected in different parts of the organisation. Effective desensitisation is when sensitive elements of a data set are anonymised, while maintaining the analytical value of the data.

As organisations acquire increasing volumes of personal and other confidential data, and interact with it in increasingly complex ways to derive value, the need for more sophisticated tools to ensure the preservation of confidentiality becomes more acute.   In order to explore this particular area, the Bank chose to work with Privitar.

Privitar 'Publisher'

Privitar's 'Publisher' tool is a software application which takes sensitive data as input and de-identifies it to create a 'safe' copy suitable for analytics, machine learning or sharing with third parties. Confidentiality is preserved by redacting, tokenising, blurring or statistically generalising certain values in the records, in order to mitigate the risk of any individual being identified, or of sensitive information about an individual being deduced from the data. The software therefore limits the possibility of reverse engineering of data via linkage attacks, whilst preserving the analytical utility of the data.

The software performs anonymisation according to centrally managed and audited policies, managed through a central web application, in order to allow transparency, consistency and accountability in how anonymisation is performed.

The Proof of Concept

The Bank undertook a Proof of Concept (POC) with Privitar to further improve its understanding of the tools and technologies used in desensitisation, and to what extent this can play a role in enabling the safe sharing of data – whether that be more widely across institutions, or potentially, with external research partners.

The Bank provided the firm with a synthetic dataset of four million mortgage records with the aim of exploring to what extent, data continued to retain its analytical value, once it has been de-identified to protect against attacks.

Reflections and next steps

The Bank believes there is much to be gained from unlocking the safe sharing of data. The initial insights gained from this PoC with Privitar have helped to showcase some of the emerging privacy engineering techniques that organisations will need to draw on to use customer data to innovate, whilst protecting sensitive and personal information.