DataSecTech

SecureCompute for Insurance Fraud Detection

DST SecureCompute

Context

In the United States alone, the total cost of non-health insurance fraud is estimated to be more than $40 billion per year, costing a hundreds of dollars increase in annual premiums for legitimate users . For example, there may not be a not simple way to check whether the same claim is filed with multiple companies. In many cases, an increased scale of shared data could lead to improved predictions and analysis for detecting insurance fraud. For example, an increased scale of claims data combined with other unstructured data will allow institutions to better identify potentially fraudulent insurance claims.

Privacy-preserving Data Sharing, Linkage and Analytics Opportunity

As a starting point, the data from multiple insurance providers could be linked to identify duplicate claims, filed against the same assets or the same incident across multiple insurers. However, much of this claim data is privacy sensitive (e.g., registration data, claims data, personal information, third party reports) and may not be easily shared among insurance companies due to privacy concerns. Furthermore, insurance companies may not be willing to share such information with their competitors, since it could be leveraged to infer commercially sensitive information such as underwriting and pricing strategies. Therefore, using a privacy preserving solution, insurance companies can link their data to improve their fraud detection models without disclosing sensitive corporate or privacy sensitive customer information.

Our solution

At DataSecTech, we provide a cloud based secure record linkage and analytics solution that could be used to address this important use case. Using our solution, different insurance companies can link their insurance claim data even if they do not have a unique identifier such as social security number or records that have errors (e.g., typos) to check whether duplicate claims have been filed.

For linkage purposes, users of our tool can choose different attributes (e.g., names, surnames address etc) and linkage algorithms. This way, insurance companies may check whether the same address is used for the same type of claim around the similar time frame. Using our privacy-preserving solution, the linkage accuracy would be as close as possible to state-of-the-art record linkage solutions that provide no privacy protection.

To provide security and privacy, the shared claim data that is used for linkage and analytics purposes is sanitized and encrypted. Using the existing confidential computing cloud and hardware support available, the entire process is end-to-end encrypted and the plaintext data will not be accessible to us and/or the cloud provider.

Once the records are linked, our solution allow further processing such as outputting required statistics, learn a machine learning model (e.g., learn a new model for predicting emerging insurance fraud patterns), or apply a given ML model on the linked dataset (e.g., use a previously developed model to flag potential insurance frauds).

Finally all the required results will be sent back encrypted to the users.

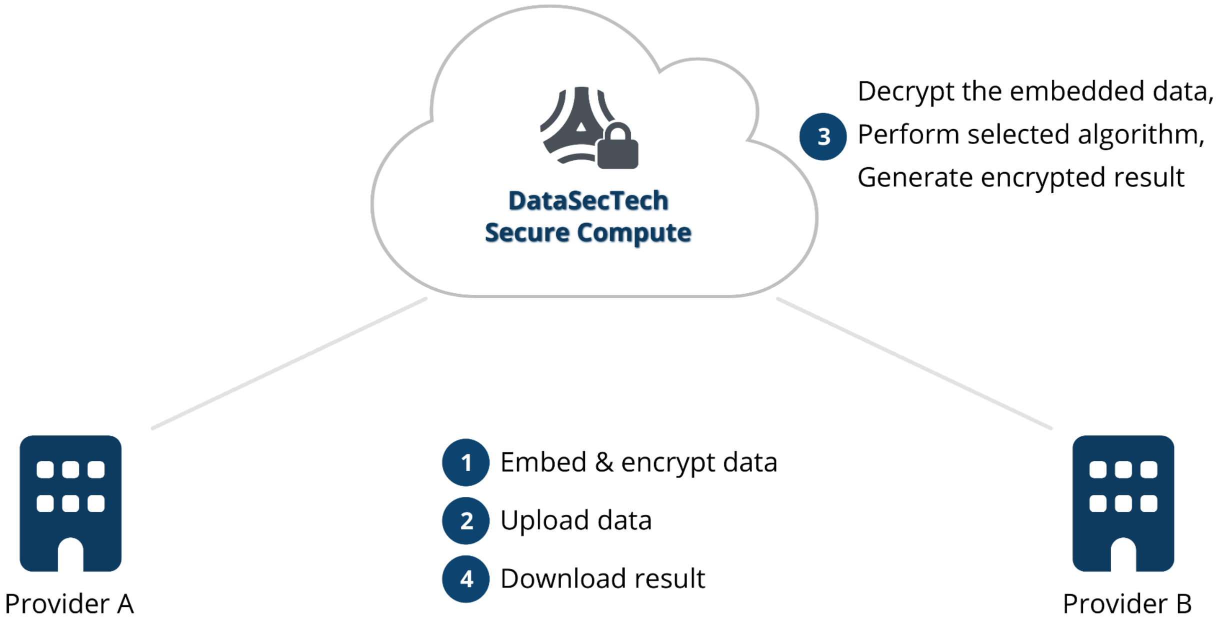

Figure 1: Overview of DST Secure Compute service

Benefits

1. Cloud based and deployable on existing cloud infrastructure
2. Entire data is sanitized encrypted end-to-end.
3. Protects privacy and complies with existing privacy regulations.
4. Resilient to errors and typos in the attributes such as names, addresses, date of births etc. used for linking.
5. Further processing is feasible on the linked data including learning ML models, testing the linked data with existing ML models, and getting useful statistics based on the linked data.
6. More efficient than heavy crypto solutions such as secure multi-party computation.