

Unlocking the Power of Customer Interaction Analytics



Introduction

ABN AMRO is a leading Dutch bank, providing retail, corporate and private banking services to almost 7 million customers.

As one of the largest banks in the Netherlands, with offices throughout Europe, ABN AMRO collects large volumes of unstructured data from customer interaction channels including online chat logs, social media posts and call centre transcripts.

Performing analytics on this unstructured data offers ABN AMRO a wealth of opportunities to gain valuable Voice of the Customer (VOC) insights and make data-driven decisions to improve customer experience and drive operational efficiencies. However, use of this data is currently restricted due to privacy concerns.

Privacy is paramount for ABN AMRO. To utilize unstructured data and create value for their customers safely and ethically, the bank sought to leverage privacy enhancing technologies to redact and protect the identifying information contained within unstructured data sources.

Privitar Labs partnered with ABN AMRO to investigate how innovations in privacy enhancing technologies could meet the bank's need to protect identifying information in unstructured text.



Addressing privacy risks to reach invaluable insights

Customer interaction data provides the chance for ABN AMRO to understand customers, their behaviour, and their attitudes toward the bank and its services.

Analyzing this data helps the bank to create customer-centric products and services, optimize operations and ensure full regulatory compliance. For example, understanding why customers contact the bank creates opportunities to optimize customer services and the contact channels themselves.

Provisioning unstructured data is currently restricted. The privacy risks in sharing this data are considerable given the sensitive and identifying information it might contain; customers share data such as names, addresses, account numbers and social media identities.

Regulations such as GDPR place increasingly higher standards on the bank for how customer data is safeguarded. Privacy concerns therefore currently limit ABN AMRO's use of unstructured data for analytics and machine learning.

The ability to detect and redact identifying information in unstructured data removes privacy barriers and opens up the value of this

data for ABN AMRO through analytics. Redaction allows unstructured data to be provisioned to a wider group of analysts while upholding customer privacy and ensuring regulatory compliance.

The research collaboration between ABN AMRO and Privitar Labs focussed on unstructured customer interaction data in the form of online chat logs and social media messages. ABN AMRO would use this data for VOC analysis, including sentiment analysis and intent labelling. ABN AMRO requires enterprise-grade data redaction capabilities to provision safe, high-utility data from these unstructured data sources.

ABN AMRO performs text analytics tasks using IBM Watson installed in a dedicated secure environment separate from the unstructured text data sources. The ability to detect and remove sensitive information before the data is provisioned to the Watson environment would reduce the bank's exposure to compliance risk and enable the data to be used by a wider set of teams.



Innovative detection for unstructured text

Privitar's software prototype uses Named Entity Recognition (NER) machine learning models to detect identifying information in free text.

Prior to engaging with ABN AMRO the prototype was achieving state-of-the-art accuracy for several common identifiers, including name, location, social media ID and URL. However, when deployed on a new data source containing new entities not previously seen by the machine learning model during training, new entities were not detected.

Identifying entities are terms within text which can include personal identifying information (PII) and protected characteristics. Name, location, phone number and social media ID are common types of identifying entities which may be found in unstructured text. The goal was to teach the NER model to efficiently learn new

identifying entities not previously detected, while maintaining the high levels of accuracy.

Training the NER model to efficiently learn new entities would enable ABN AMRO to onboard new use cases with minimal impact on business operations. However, large quantities of manually labelled records are needed to train a machine learning model to learn a new entity. Our goal was to reduce the number of labelled records required – to learn more with less data. This would create significant time and resource efficiencies, with staff focussed on operational delivery rather than data labelling.

Unlocking the value of unstructured data

Independent workstreams at ABN AMRO and Privitar Labs had initial success developing techniques to de-identify unstructured text and achieving high accuracies detecting a known set of entities in test datasets.

Training deep learning models used for NER typically requires large amounts of labelled data. This involves humans manually labelling thousands of example sentences containing the new entity to be learned. Reducing the amount of time and resources to label data is a key business requirement for ABN AMRO to mitigate the high costs of onboarding individual use cases. A solution

that required only small quantities of data to be labelled would help ABN to drive efficiencies, onboarding new use cases quickly and keeping their staff engaged on operational activities.

As a banking organization, detecting currency amounts was a key requirement for ABN AMRO. However, currency amounts appear relatively infrequently in datasets. This made it even more important to develop a technique to efficiently identify records in a dataset without the need for a human to sift through high volumes of data to find examples.

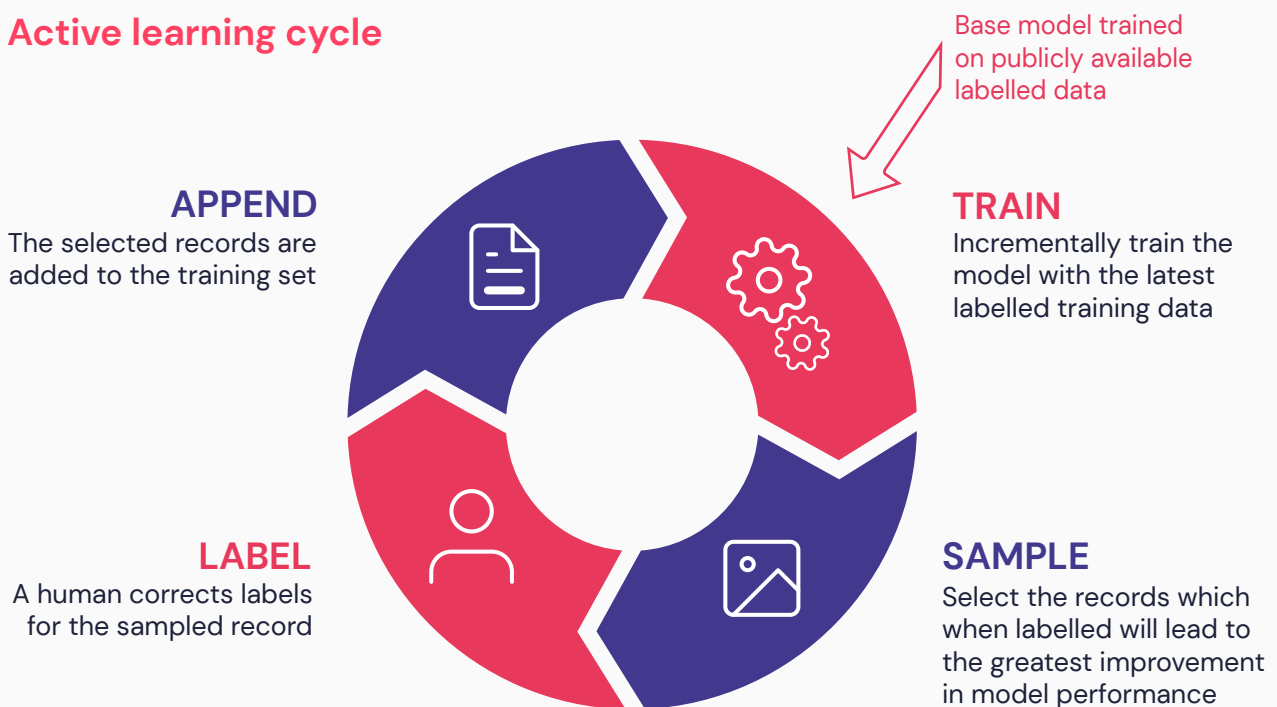


Special consideration was given to ensuring the solution delivered high accuracy to ensure preservation of customer privacy. ABN AMRO and Privitar set an aggressive target of training the NER model to achieve 80% accuracy on the new

currency entity with less than 500 training records. This is equivalent to 1 hour of human labelling time. It was also paramount that there would be no impact on the ability to accurately detect existing entities.

Figure 1 shows the active learning cycle used in training the NER model to efficiently learn currency as a new entity.

Active learning cycle



Active learning delivers impressive results

Working in partnership with ABN AMRO, Privitar developed a solution that trains the NER model to detect new identifying entities without impacting accuracy for existing entities.

Real customer data was used to verify this. Privitar used a technique called active learning – which combines human expertise and machine efficiency – to meet ABN AMRO's need to reduce human time spent on labelling records.

The active learning sampler seeks out the most valuable records for a human to label. These records are used in turn to train the NER model in an iterative cycle of improvement to drive up accuracy. For reliable accuracy, humans still need



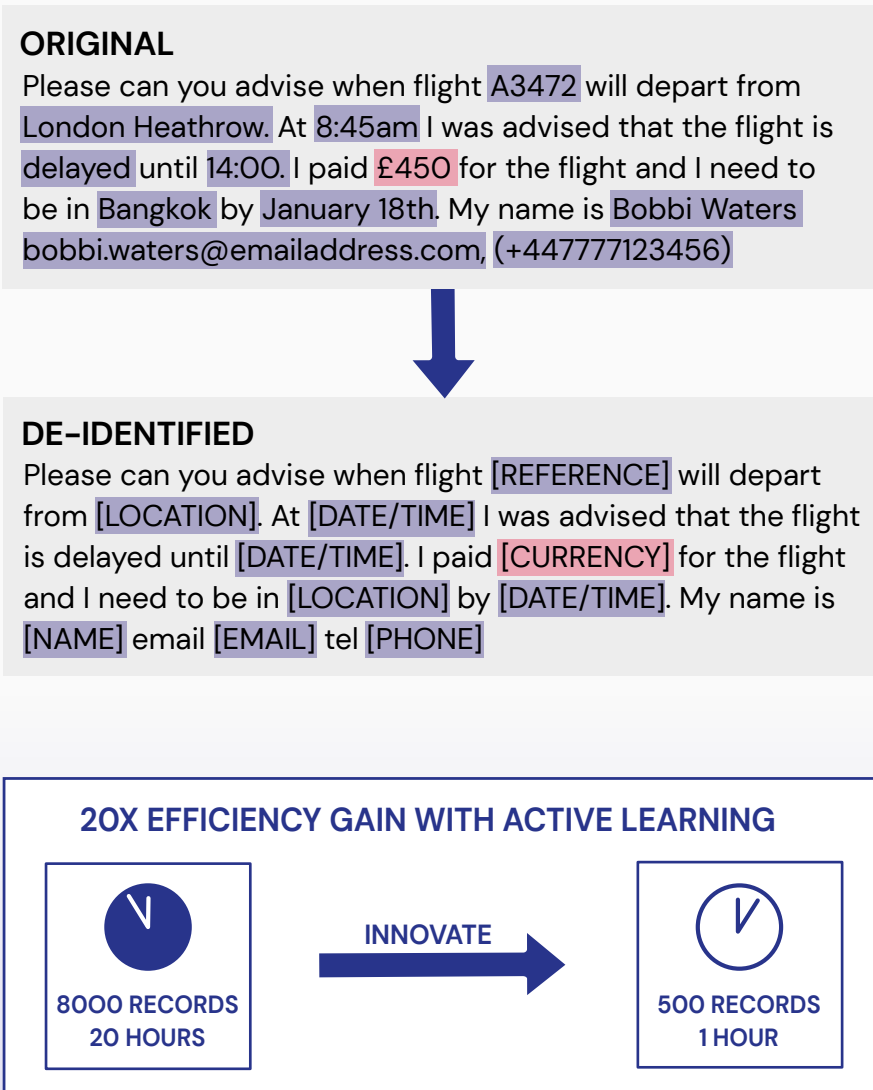
to label example records, bringing their ability to understand nuances in the use of language. The active learning technique makes the most efficient use of human effort in teaching the NER model to learn a new entity.

Using the active learning technique, ABN AMRO's requirement to learn currency as a new entity with a minimum accuracy of 80% was achieved using just 5% of the volume of data that would have previously been required. If ABN AMRO

uses this technique to onboard a new identifying entity, the amount of time needed from a human experienced at labelling will be reduced from twenty hours to just one hour.

ABN AMRO was also able to deploy Privitar's prototype on Dutch customer data and evaluate its efficacy in learning to identify URLs in Dutch language chat logs and WhatsApp data, achieving the same efficiency gain as with English language examples.

Figure 2 shows a sample of unstructured text pre- and post-redaction by the Privitar prototype, and the operational efficiencies which are gained through training the model with active learning.





Conclusion

Privitar's prototype for redacting identifying information within unstructured data can be trained to detect new entities specific to individual customer requirements, in this case meeting ABN AMRO's requirements to learn currency as a new entity. The collaboration also directly shaped and improved the prototype functionality.

Privitar's prototype unlocked access to customer interaction data for ABN AMRO, enabling the bank to provision unstructured data for analytics, safely. The prototype exceeded ABN AMRO's requirement

to reduce the amount of time required by a human to label data in order to train a model for machine learning. Privitar's active learning strategy generated twentyfold gains in operational efficiencies for manual labelling.

The capability demonstrated in this Privitar Labs project has the potential to open up the value of unstructured data for ABN AMRO across the whole organization, with machine learning models able to learn new entities efficiently and without restriction in greatly accelerated timescales.

About Privitar Labs

Developing advanced solutions to the most challenging data privacy problems.

The practical application of Privacy Enhancing Technologies (PETs) is advancing rapidly. Many promising new techniques are emerging, but the challenge lies in matching these to the right business use cases, and developing them into products that are easy to understand and to use correctly, at scale. This is Privitar's mission.

Within Privitar Labs, we're driving the creation of practical solutions using PETs. We work in close partnership with our strategic customers to apply leading privacy techniques to enable new uses of data.

Privitar is the leader in modern data provisioning, empowering organizations to maximize business value by using data safely, with speed and at scale.

We make data highly accessible, through the use of privacy enhancing technologies, so organizations can optimize business and customer outcomes. Only Privitar has the right combination of technology and expertise to create a safe data provisioning ecosystem to enable clients to share data and unlock new data insights while keeping data safe and businesses compliant. Founded in 2014, Privitar is headquartered in London, with regional headquarters in Boston and locations throughout the US and Europe.

For more information, please visit www.privitar.com



[PRIVITAR.COM](https://www.privitar.com)