WHITE PAPER

Como IA e Machine Learning Estão Redefinindo Práticas Anti-Lavagem de Dinheiro

Os criminosos estão atacando com força. Use toda e qualquer vantagem para combatê-los.

sas

Conteúdo

O panorama em evolução da lavagem de dinheiro

Criminosos financeiros são bons em esconder a origem de seus ganhos ilegais e em colocar esse dinheiro de volta no sistema financeiro, tanto para ganhos pessoais quanto para financiar outros crimes. Enquanto a maior parte da lavagem de dinheiro provém do tráfico de drogas e crime organizado, os eventos de 11 de setembro de 2001 também colocam o foco no financiamento secreto de atividades terroristas, que é tradicionalmente mais difícil de detectar.

É desafiador. Considere o imenso volume de dados que é esperado que as instituições financeiras consigam peneirar para atender às regulações e detectar e reportar atividades suspeitas. Os dados são normalmente variados e abaixo da média. É comum que sistemas usem apenas um conjunto de dados disponíveis ao gerar alertas. Sistemas tradicionais de monitoramento de transações são difíceis de manter e dependem de regras e limites fáceis de serem testados e contornados por criminosos. Processos de investigação tendem a ser altamente manuais, desde a coleta de informações para um caso até o envio de um Relatório de Atividade Suspeita (SAR, na sigla em inglês) completo.

Enquanto isso, os criminosos continuam trabalhando noite e dia para permanecerem escondidos, constantemente inventando novas maneiras de ocultar o fluxo de fundos.

As ferramentas e táticas anti-lavagem de dinheiro e de combate ao financiamento do terrorismo (AML e CFT, respectivamente na sigla em inglês) levam mais tempo e dinheiro do que deveriam. Para fortalecer a defesa, as instituições financeiras precisam de maneiras de:

- **Automatizar tarefas** que precisam formalmente de intervenção humana, como a disposição de alertas**.**
- **Detectar melhor riscos** e priorizá-los efetivamente com técnicas de inteligência analítica sofisticadas**.**
- **Fornecer contextos melhores** para as investigações com acesso a insights mais abrangentes**.**

É aqui que a IA entra em cena

O conceito de inteligência artificial (IA) evoca visões de robôs que aprendem demais, concedem a si mesmos muito poder e eliminam seus criadores. A realidade da IA é muito menos dramática. De modo geral, é permitir que uma máquina tome uma decisão que um humano poderia ter tomado.

Quando a Amazon e a Netflix recomendam coisas que você pode gostar, IA está por trás disso. Quando as assistentes virtuais Siri e Alexa ajudam a organizar sua vida e fazem recomendações ou quando o reconhecimento facial realiza autenticações ou pagamentos online, isso é a IA em ação. De carros autônomos a ofertas personalizadas à detecção de fraude de cartão, as tecnologias de IA e machine learning estão beneficiando uma série de indústrias e criando mercados que nunca imaginamos.

.

Um subconjunto de IA, machine learning habilita um programa de computador a aprender com dados, ao invés de através de uma programação explícita. Esses programas funcionam utilizando amostras de dados, encontrando padrões neles que possam ser complexos demais para um ser humano ver intuitivamente e depois aplicar essas descobertas a novos dados. Ao unir essa capacidade de aprendizagem com o poder computacional moderno, você tem uma receita para um sistema que pode tomar decisões complexas de maneira automatizada.

Algumas técnicas de machine learning podem extrair novos elementos de dados desconhecidos ou indisponíveis previamente para modelos AML ou CFT. Esses modelos podem interpretar padrões previamente desconhecidos de grandes fontes de dados, que podem vir a alimentar outros esforços em modelagem. Outras técnicas de machine learning pdem ser usadas para fazer predições diretas baseadas nos padrões encontrados. Quantos mais treinamento um modelo tem com dados de feedback, mais preciso ele se torna e menos ajustes são necessários.

As faces de machine learning

A abordagem certa de machine learning para AML/CFT depende das informações que você tem para treinar o modelo e o que você espera alcançar.

Machine learning supervisionado

Com um machine learning supervisionado, são inseridos dados de entradas amostrais e suas repostas associadas no modelo com o objetivo de gerar uma regra geral que mapeie essas entradas e saídas. Por exemplo, quais atributos foram associados com casos que acabaram sendo SARS? Quais descobertas foram associadas com falso-positivos ou falso-negativos? O modelo aprende a prever melhor o resultado quando é aplicado a dados novos. Técnicas de aprendizagem supervisionada tradicionais incluem:

- **Estatísticas Bayesianas,** que descrevem a probabilidade condicional de um evento baseado em entradas de dados, assim como informações anteriores ou teorias sobre o evento.
- **Árvores de decisão,** um modelo de decisões "se/então"; e suas possíveis consequências.
- **Redes neurais,** onde "neurônios" são conectados por grau de importância, uma técnica amplamente utilizada para reconhecimento de voz, análise de imagens e controle adaptativo.
- **Análise de regressão** para modelar e analisar as relações entre variáveis dependentes e uma ou mais variantes independentes.
- **Floresta aleatória,** um método de aprendizagem em conjunto para classificação, regressão e outras tarefas, que funciona construindo uma grande quantidade de árvores de decisão.

Ao observar conjuntamente todas as contas, atributos de cliente e eventos, soluções de machine learning não supervisionado podem revelar padrões significativos que não seriam vistos com sistemas AML tradicionais baseados em dados.

Machine learning não supervisionado

Com machine learning sem supervisão, o algorítimo aprende a partir de dados amostrais que não tenham sido rotulados, classificados ou categorizados.O algoritmo precisa encontrar sozinho padrões ou estruturas ocultas nos dados. Uma vez que você não sabe quais dados representam atividades suspeitas, você precisa que o modelo

crie uma função que descreva a estrutura dos dados, aponte anomalias e, então, aplique esse conhecimento aos novos dados.

Técnicas comuns de aprendizagem não supervisionada incluem:

- **Análise de afinidade**, uma técnica de análise e mineração de dados que revela relações entre atividades realizadas por (ou gravadas sobre) entidades específicas.

- **Clustering**, uma forma de mineração de dados exploratória que agrupa "objetos" em um grupo (cluster) onde esses objetos são mais similares entre si de alguma forma do que em relação aos de outro cluster.
- **Mapeamento KNN (nearest neighbor)**, um exercício de otimização que consiste em encontrar um ponto em um dado conjunto que esteja mais próximo (ou seja mais parecido) com qualquer ponto dado.

Resumidamente, machine learning permite que um programa de computador aprenda com os dados, ao invés de através de uma programação explícita. O programa usa dados amostrais, encontra padrões que podem ser muito complexos para seres humanos enxergarem e, então, aplica as descobertas a novos dados.

Com os avanços no poder computacional moderno, decisões altamente complexas podem ser automatizadas com grande velocidade e precisão. É fácil enxergar o valor de machine learning para a manutenção do ritmo dos esquemas em evolução de AML/CFT.

Modelos podem processar enormes quantidades de dados demográficos, geográficos e de informações sobre atividades dos criminosos e suas redes de contato. Armados com uma visão holística e global, investigadores podem executar pesquisas mais eficazes.

Machine learning em funcionamento – seis casos de uso para AML

Graças a avanços no tratamento de big data, machine learning agora pode mudar a arquitetura de AML. As instituições financeiras podem:

- Trocar seus mecanismos baseados em regras por modelos de machine learning.
- Usar machine learning como sistema de suporte para desenvolver modelos que alimentam seus mecanismos tradicionais, trazendo nova inteligência para atividades como classificação de risco, ajuste de regras e priorização de alertas.

Vamos ver algumas das maneiras que machine learning já está sendo colocada em uso para AML.

Machine learning como complemento ao monitoramento de transações

Alertas históricos associados a comportamentos suspeitos, disposições produtivas ou arquivos de relatórios de atividades suspeitas – alertas conhecidos como "bons" – podem ser rotulados e um modelo supervisionado de machine learning pode ser treinado com esses dados para se tornar mais eficiente em pontuar e priorizar novos alertas. Os investigadores podem, então:

- Focar nos alertas com maior probabilidade de resultar em arquivos de relatórios de atividades suspeitas.

- Revelar ligações ocultas entre os criminosos de hoje e seu histórico de parceiros.

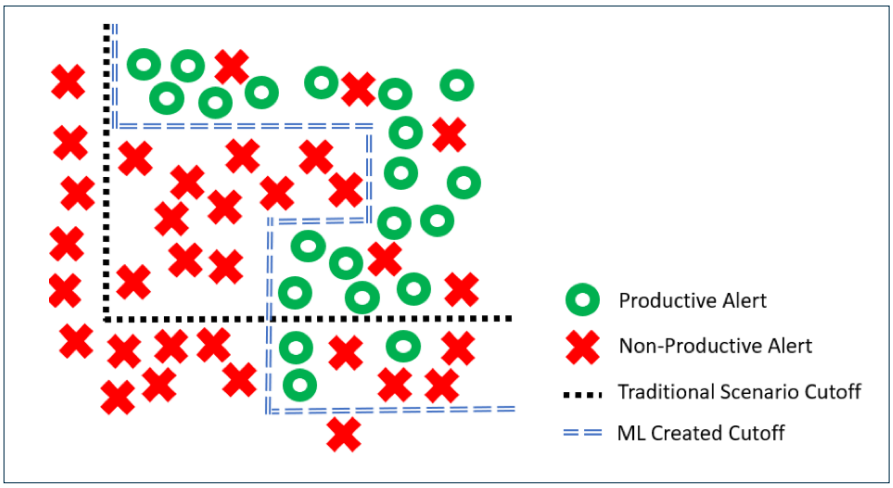

Figura 1. Um modelo machine learning pode fornecer mais alertas produtivos de atividades suspeitas ao considerar relações mais complexas.

Uma possível abordagem é projetar um modelo machine learning de priorização de alertas em um sistema existente de monitoramento de transações para classificar alertas por seus níveis de desconfiança. Ou um modelo machine learning pode ser projetado para aumentar ou substituir completamente cenários baseados em regras. De qualquer modo, um modelo machine learning pode ajudar investigadores a priorizar filas de fluxo de trabalho para revisão ou a auto disposição de alertas – escalonar alertas de alto risco e "hibernar" os de baixo risco.

Com modelos machine learning você pode:

- **Reduzir falso-positivos com uma segmentação mais firme.** Cenários baseados em regras e modelos de segmentação normalmente são amplos e baseados em dados convencionais, como demográficos ou de tipos de contas. Uma abordagem ampla leva a mais falso-positivos, mais carga de trabalho para investigadores e um custo mais alto para a organização. Em contraste, técnicas de machine learning, como clustering, podem definir segmentos mais refinados para uma avaliação de risco mais focada.
- **Aumentar positivos verdadeiros com inteligência analítica poderosa.** Modelos machine learning podem identificar atividades e comportamentos suspeitos que sistemas de monitoramento tradicionais, baseados em regras, não veriam. Esses modelos podem revelar ligações e relações ocultas , revelando novos padrões de atividades suspeitas que foram previamente indetectados.

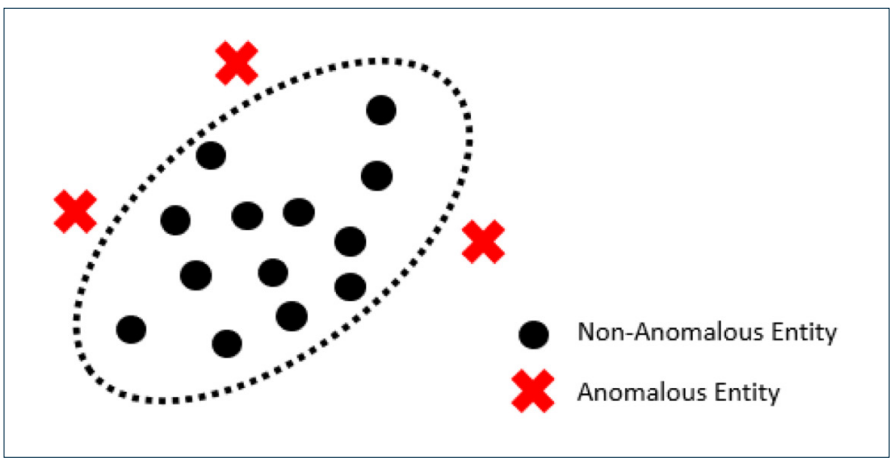

Figura 2. Um exemplo simples e conceitual da detecção de pontos fora da curva baseado em duas variáveis.

Um banco da região Ásia-Pacífico reduziu falso-positivos em 33% ao desenvolver uma classificação de alertas e abordagem de "hibernação" para prever investigações que valessem a pena.

Um banco americano de capital tier 2 modernizou seu sistema AML de monitoramento de transações baseado em regras ao implantar um modelo de redes neurais SAS®, que reduziu itens de trabalho em 50% e aumentou a taxa de conversão SAR de 5 para 15%.

Machine learning na detecção de anomalias

Programas AML são baseados no conceito de encontrar atividades suspeitas, mas não há uma definição objetiva de "suspeita". Isso é um problema para machine learning supervisionado, que precisa aprender com exemplos rotulados como "suspeito" ou "não suspeito".

Técnicas de detecção de anomalias resolvem esse problema ao identificar observações que aparentam ser matematicamente "distantes" do esperado – diferentes do seu próprio histórico de atividades ou da atividade atual de colegas. Os pontos fora da curva – comportamentos fora da norma – garantem um olhar mais atento. Com pouco direcionamento e sem dados classificados, esses métodos de detecção de anomalias podem encontrar atividades potencialmente suspeitas não definidas por uma regra.

Machine learning para a segmentação de clientes

Clientes diferentes realizam transações de maneiras diferentes e, portanto, devem ser monitorados diferentemente. Portanto, naturalmente, programas AML segmentam clientes em grupos de entidades semelhantes. Há muitas maneiras lógicas de segmentar grupos assim – por demografia, estrutura de negócios ou volume de transações, por exemplo. Mas algoritmos de machine learning não supervisionados oferecem possibilidades mais sofisticadas.

Por exemplo, um algoritmo de clustering como K-means clustering classifica entidades em grupos similares ao identificar observações que são matematicamente "próximas" entre si por um ou múltiplos fatores. Assim como duas cidades podem ser semelhantes baseadas em uma combinação de fatores como distância física, densidade e tamanho populacional, diferentes entidades podem ser vistas como similares entre uma combinação de atributos e serem monitoradas de acordo.

Machine learning para classificação de risco de cliente

Algoritmos de machine learning podem usar dados conhecidos sobre o histórico de clientes e atividades suspeitas que resultam em arquivos de relatórios de atividades suspeitas para classificação de risco (ou pontuação) de outros clientes com atributos similares. Essa abordagem pode ser combinada com dados de fora da organização para oferecer uma visão holística e global do cliente como parte do processo de conhecer seu cliente. Modelos de machine learning podem ser treinados para triagem e integração de clientes ao atribuir pontuações e usá-las como rótulos para treinar os modelos de sistemas de monitoramento de transações.

Machine learning para análise de redes sociais

Machine learning pode identificar indivíduos e redes entrelaçadas de atividades suspeitas. Modelos de machine learning não supervisionados podem verificar contas e clientes envolvidos em transações com malfeitores ou pessoas na lista de sanções de pessoas politicamente expostas (PPE). Modelos de machine learning podem revelar rapidamente padrões ocultos ligações e conexões entre milhões de contas e variáveis – e conectar criminosos com outros indivíduos da sua rede de contatos.

Machine learning para definição e ajuste de limites

Para ser eficaz, modelos e regras precisam se adaptar ao que está acontecendo no

Um banco global de capital tier 1 testou a validade de aplicar IA para detecção de riscos ocultos ou falso-negativos dentro da sua base de clientes e identificou 416 clientes suspeitos de operar como empresa de serviços financeiros, 89 dos quais não haviam sido detectados previamente.

Um banco global de capital tier 1 aprimorou sua experiência de cliente para negócios com financeiras comerciais ao acelerar os tempos para diligência prévia e aumentando a precisão. O banco implantou algoritmos deep learning automatizados para classificar os tipos de documentos sob revisão, reduzindo assim os esforços no processo de revisão de duas pessoas por duas semanas para cerca de um minuto.

mundo. Imagine o poder de um sistema que pode automaticamente examinar grandes volumes de dados para ajudar a estabelecer regras e mantê-las atualizadas. Machine learning pode ser usada para estabelecer limites derivados de dados que calibram a si mesmos baseados no que aprendem. Esses modelos podem processar grandes volumes de dados em alertas produtivos para estabelecer limites mais exatos e precisos. Baseado no resultado desses modelos, as empresas podem realizar ajustes automáticos contínuos – uma maneira melhor de gerenciar um processo que, de outro modo, seria muito demorado.

Os três principais desafios da adoção de machine learning

Apesar dos benefícios comprovados de machine learning em múltiplos mercados, instituições financeiras estão demorando a adotá-lo para AML. A partir da experiência SAS com clientes nesse mercado, três problemas são os maiores empecilhos:

- Faltam o conhecimento, as pessoas e os sistemas necessários.
- Preparar dados é difícil.
- Modelos de machine learning tendem a ser sistemas "caixa preta".

Vamos olhar esses desafios mais atentamente.

Faltam o conhecimento, as pessoas e os sistemas necessários

Não é segredo que há uma escassez de talentos profissionais em ciência de dados. Apesar do surgimento de muitas novas graduações e programas de certificação, as empresas ainda lutam para encontrar e manter profissionais qualificados para atender suas ambições analíticas. Um bom processo de desenvolvimento de modelo é repetitivo, contínuo e demorado. Isso implica em realizar testes em vários tipos de modelos, combinações de entrada de variáveis e parâmetros de modelos – e então, continuamente atualizá-los, testá-los e reimplantá-los.

Para organizações sentindo falta de talento em ciência, parte da solução é ter um conjunto de ferramentas de software que automatize muito desse trabalho analítico e gerencie o ciclo de vida do modelo. De qualquer modo, ainda é muito útil ter a experiência humana à mão para projetar e ajustar esses sistemas de maneira inteligente e eficaz.

O SAS tem uma grande seleção de consultores altamente experientes que implementam soluções em uma grande variedade de mercados. Nosso time de Soluções de Segurança Inteligente consiste em especialistas de mercado com experiência em análises AML em instituições bancárias de capital tier 1.

Preparar dados é difícil

Para um algoritmo de machine learning funcionar, os dados para projetá-lo precisa ser identificado, coletado, fundido a um data mart centralizado e gerenciado. Até mesmo dados supostamente "limpos" podem ter problemas desconhecidos como pontos fora da curva ou áreas ausentes.

Modelos machine learning supervisionados também requerem resultados amostrais – por exemplo, dados sobre alertas produtivos e não produtivos. Algumas empresas podem ser capazes de usar dados de um sistema existente de AML, mas qualquer nova característica precisará de dados recentemente classificados, o que exige trabalho dos investigadores.

Essas etapas necessárias de preparação de dados são, frequentemente, negligenciadas.

O SAS colocou arquiteturas e processos de dados em funcionamento para atacar esses problemas de gestão de dados como parte da nossa solução padrão de AML.

Modelos de machine learning tendem a ser sistemas caixa preta

Enquanto modelos de machine learning são excelentes em encontrar padrões, eles não o fazem de maneira fácil de entender e explicar a um regulador. Muitos dos modelos de machine learning que recebem muita fama atualmente – redes neurais, floresta aleatória e gradient boosting – são abordagens opacas. Os resultados desses modelos é complexo ao ponto de ser impraticável decifrar. Existem técnicas para, através de engenharia reversa, decifrar o resultado de maneira a entendê-lo melhor, mas esses modelos ainda são desafiadores para explicar.

De modo alternativo, técnicas de modelagem estatística "caixa branca", como modelos de regressão e árvores de decisão, trocam um pouco de poder preditivo por transparência, o que é útil para explicar a um regulador. E cenários baseados em regras são naturalmente o mais fácil para entender e explicar.

Então, que tipo de inteligência analítica sua organização deveria usar? Depende da natureza do seu problema e o nível de opacidade de modelo que sua empresa aceite. Instituições financeiras mais analiticamente maduras tem visto resultados positivos com modelos de caixa preta, mas nem toda organização está pronta para dar esse passo. Dependendo dos dados e uso de caso, o ganho incremental de uma abordagem mais complexa pode não ser justificado. Você pode decidir experimentar ambas para obter uma compreensão completa dos prós e contras.

O SAS apoia ambas as abordagens e pode ajudar a guiá-lo por essa decisão.

Como o SAS Pode Ajudar

O SAS vem desenvolvendo software de inteligência analítica há mais de 40 anos. Trabalhando junto com clientes de serviços financeiros, nós desenvolvemos uma solução que atende processos AML através de áreas importantes, como o monitoramento de atividades suspeitas, diligência prévia de clientes, filtragem de listas de observação e gestão de casos de investigação.

Principais recursos

- **Gestão de dados.** Resolva os desafios de dados AML, do processamento de big data ao acesso e integração de fontes legadas – tudo em uma única plataforma.
- **Visualização e inteligência analítica de alto desempenho.** Obtenha insights rápidos a partir de big data com uma infraestrutura que te permite testar hipóteses, fazer perguntas e simular cenários.
- **Relatórios e Monitoramento de atividade suspeita.** Conte com um mecanismo de cenário flexível e robusto que identifica com mais precisão atividades suspeitas e gera alertas para eventos que atendem a parâmetros definidos.
- **Correspondência de lista de observação.** Utilize algoritmos de correspondência difusa, classificação inteligente e consolidação de alertas para identificar pessoas, organizações ou jurisdições que representem riscos regulatórios.
- **Gestão de investigação e alertas.** Use uma interface baseada na web para obter uma visão completa de itens de trabalho investigativo, com fácil acesso a uma base de dados central de conhecimento.
- **Detecção de anomalias em grupos de entidades pares.** Inteligência analítica in-memory identifica rapidamente atividades potencialmente suspeitas ao comparar o comportamento de uma entidade a comportamentos históricos e de seus pares.
- **Busca.** Pesquise grandes repositórios de dados com replicação e indexação distribuída e consultas load-balanced, failover automático e recuperação, entre outros.
- **Arquitetura Multitenant.** Atenda múltiplos grupos com segurança com uma instalação única ao segregar os dados.

A mais nova geração de software de machine learning do SAS® é oferecida na plataforma SAS® Viya®, que funciona em um ambiente distribuído e in-memory. o SAS automatiza os processos associados com AML de ponta a ponta, de gestão de dados e gestão de modelos a gestão de casos e governança.

SAS Viya atende a uma variedade de usuários, de analistas de negócios a cientistas de dados. Há interfaces intuitivas "arrasta e clica" ou a opção de programar com a linguagem de programação SAS ou APIs de suporte de outras linguagens de programação populares como R e Python. E, é claro, há visualizações de dados intuitivas que permitem que usuários de negócios explorem os dados em diversas dimensões. Seus desenvolvedores, analistas e líderes podem escolher seus meios de trabalho de preferência.

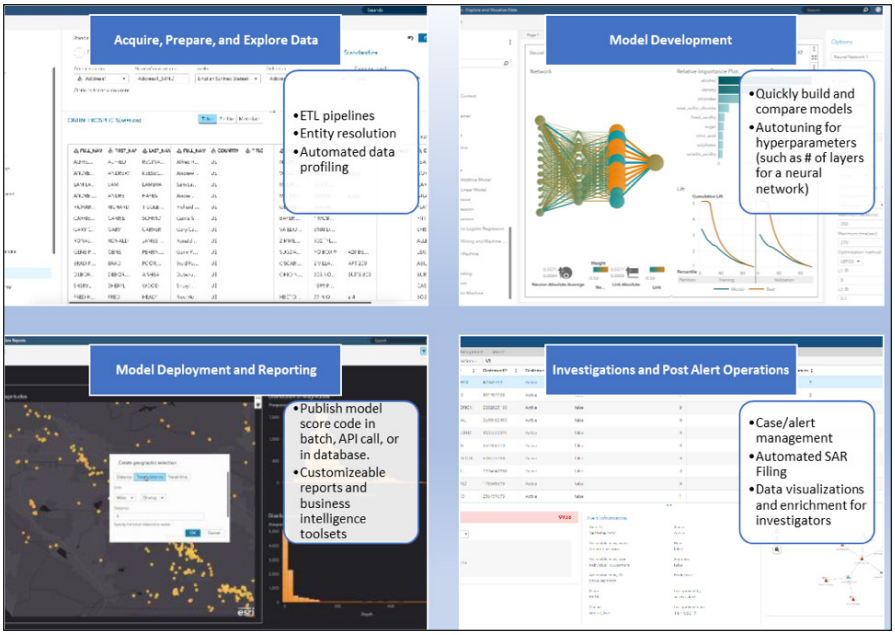

Figura 3. O SAS oferece suporte ao continuum AML, desde a gestão de dados e modelos até a gestão de alertas, investigações e relatórios.

Saiba mais em sas.com/en_us/software/anti-money-laundering.html.

Entre em contato conosco em GHUSPSDAMLPresales@sas.com.

Para entrar em contato com o escritório local SAS, visite: sas.com/offices