Detecção de Fraudes na Emissão de Certificados Digitais dentro da Infraestrutura de Chaves Públicas Brasileira

Fernanda O. Gomes[1] **Bruno M. Agostinho**[1] **,**
Julia Baldisera[1] **, Raphael S. da Silveira** [1] **,**
Jean E. Martina[1]

[1]Universidade Federal de Santa Catarina
Departamento de Informática e Estatística
Campus Universitário – Florianópolis – SC – Brasil

fernanda.gomes@posgrad.ufsc.br, bruno.agostinho@posgrad.ufsc.br,

juhbaldisera@gmail.com, raphass22@gmail.com

jean.martina@ufsc.br

Abstract. *In Brazil, it is possible to interact virtually with any e-gov system through the use of digital certificates issued by a government-controlled Public-Key Infrastructure (ICP-Brasil). The digital certificates are a perfect tool for malicious people thief others identity virtually. Fraud using digital certificate can open access, for instance, to systems authentication and confidential documents. The current process is manual, costly, subject to human failure, and corruption. Given this, our work proposes the automation of the fraud detection process. We propose a hybrid machine learning approach using a clustering technique based on DBSCAN and classification using the data captured in the process of issuing digital certificates established by the ICP-Brasil. This fraud detection system makes the ICP-Brasil system safer, faster, and cheaper.*

Resumo. *Muitas das interações com os sistemas de governo eletrônico (e-gov) no Brasil são realizadas por meio do uso de certificados digitais emitidos por uma infraestrutura de chaves públicas controlada pelo governo (ICP-Brasil). Os certificados digitais são uma ferramenta perfeita para que pessoas se passarem por outras pessoas virtualmente, roubando assim sua identidade. A fraude na emissão de um certificado digital pode abrir portas para autenticação em sistemas e acesso a documentos sigilosos. O processo atual é manual, custoso, está sujeito a falhas e corrupção humana. Visto isso, este trabalho propõe a automatização do processo de detecção de fraudes através da proposta de uma abordagem híbrida de aprendizado de máquina utilizando uma técnica de clusterização baseada no DBSCAN e classificação usando os dados utilizados no processo de emissão de certificados digitais da ICP-Brasil. Esse sistema de detecção de fraude torna o sistema ICP-Brasil mais seguro, mais rápido e mais barato.*

1. Introdução

No Brasil, muitas das interações com os sistemas de governo eletrônico (e-gov) são realizadas por meio do uso de certificados digitais emitidos por uma infraestrutura de chaves públicas controlada pelo governo (ICP-Brasil). Com um certificado digital emitido

pela ICP-Brasil, advogados podem acessar o sistema judicial para enviar liminares eletronicamente; médicos podem assinar prescrições e validar exames complexos através de sistemas de telemedicina; empresas podem emitir notas fiscais assinadas digitalmente pelo sistema da receita federal; cidadãos comuns podem, pela Internet, abrir uma empresa, contas bancárias, preencher formulários de impostos, agendar consultas médicas, verificar seu status de assistência social e solicitar assistência social, entre outras coisas.

A ICP-Brasil é composta por mais de 20 autoridades certificadoras encarregadas de identificar corretamente as pessoas e emitir o certificado digital. As assinaturas digitais produzidas com certificados digitais são tratadas da mesma forma que as assinaturas regulares em papel pelo sistema jurídico brasileiro. Devido ao uso massivo por órgãos governamentais, os certificados digitais são uma ferramenta perfeita para que pessoas mal-intencionadas se passarem por outras pessoas on-line, roubando assim sua identidade. Fraudes desse tipo vem sendo cada vez mais divulgadas pela imprensa, sendo importante ressaltar os casos onde um juiz teve sua identidade digital fraudada para que uma sentença fosse alterada, funcionários relacionados a órgãos ambientais para conseguirem a autorização de extração ilegal de madeira e funcionários aduaneiros para evitar o pagamento de impostos de importação. Embora os números oficiais do ICP-Brasil mostrem que apenas 0,0002% dos mais de 3,5 milhões de certificados emitidos por ano são fraudulentos [ITI 2020a], os casos vêm crescendo em ritmo exponencial e sua detecção é baseada apenas na segurança processual.

O processo de emissão de certificados digitais para pessoas tem uma diferença significante comparado ao processo de emissão de certificados digitais para websites ou maquinaria. No processo atual de emissão no Brasil, o indivíduo comparecer presencialmente a um ponto de atendimento de uma autoridade certificadora com o intuito de apresentar seus documentos de identificação que serão verificados para garantir autenticidade e assim emissão do certificado. Os documentos apresentados incluem um documento de identificação (ex. passaporte, carteira de identidade, carteira de motorista), comprovante de residência, CPF, além da coleta das impressões digitais e uma foto tirada na hora.

A fraude na emissão de um certificado digital pode abrir as portas para autenticação em sistemas, acesso a documentos sigilosos, em suma a falsidade ideológica. A preocupação acerca desse tema fez com que alguns processos mais rígidos de verificação de identidade fossem adotados pelos funcionários e autoridades certificadoras. A exemplo disso, os documentos apresentados são verificados manualmente para testar sua autenticidade. São buscadas por alterações ou inconsistência nos documentos. Além disso, existe uma lista negativa de fraudadores conhecidos que já tentaram cometer fraude na emissão de certificados digitais. Os fraudadores são identificados por suas características pessoais, como idade, cor dos cabelos, cor da pele, olhos entre outras características aparentes. Esse processo de verificação é manual e pode levar cerca de 30 minutos, algo que impacta no custo final do certificado digital. Se o funcionário acreditar que pode estar enfrentando uma tentativa de fraude, ele informará o regulador do mercado que realizará uma análise forense dos documentos para avaliar se o certificado digital deve ou não ser emitido. Nos casos em que a fraude é detectada, pode levar até dois dias para que o certificado seja emitido e o custo deve ser absorvido pela autoridade de certificação e pelo indivíduo que pode ter que comparecer presencialmente várias vezes ao ponto de atendimento da autoridade certificadora. Visto isso, fica evidente à quantidade de pessoas

envolvidas no processo e como este não considera à corrupção das mesmas a até falhas humanas.

Sproule and Archer em [Sproule and Archer 2007] descrevem o roubo de identidade como um acesso não aprovado a informações ou documentos pessoais. Para operar uma fraude de identidade, um adversário usa uma identidade que não lhe pertence, o que é considerado um ato criminoso. A fraude de identidade inclui o roubo de identidade, criação de uma nova conta ou emissão de um documento usando a identidade de outra pessoa, ou roubando uma conta ou documento existente. Existem diversos trabalhos na literatura com o objetivo de detectar fraudes. A maior parte deles propõe métodos para detecção de fraudes transacionais, onde a identidade já emitida é roubada. No presente trabalho, está sendo abordado o problema de fraude em aplicações, onde um fraudador deseja emitir um documento em nome de outra pessoa, assim roubando sua identidade. Poucos trabalhos no estado da arte abordaram esse assunto [Kshirsagar and Dole 2014]. Os poucos existentes são voltados para emissão de cartão de crédito (ex. [Wheeler and Aitken 2000, Phua et al. 2009, Abdelhalim and Traoré 2010, Phua et al. 2010b]). Esses trabalhos utilizam abordagens de aprendizado de máquina supervisionado (ex. classificadores) e não supervisionado (ex. clusterização e detecção de anomalias). No entanto, nenhum dos trabalhos presentes na literatura propôs uma estrutura para detecção de fraudes de aplicação para emissão de certificados digitais.

Este trabalho tem como proposta, a automatização do processo de detecção de fraudes através de uma abordagem híbrida de aprendizado de máquina, utilizando uma técnica de clusterização semi-supervisionadas, baseada no DBSCAN, e classificação utilizando os dados capturados no processo de emissão de certificados digitais da ICP-Brasil. Esse sistema de detecção de fraude torna o sistema ICP-Brasil mais seguro, mais rápido e mais barato.

O trabalho, na sequência, apresenta alguns conceitos básicos e trabalhos relacionados nas seções 2 e 3. Na seção 4 traz o cenário atual e suas limitações. O método de detecção de fraude automatizado proposto, assim como os resultados experimentais obtidos, são apresentados nas seções 5 e 6. Por fim, a seção 7 apresenta as considerações finais com algumas observações e indicações para trabalhos futuros.

2. Conceitos Básicos

Para melhor compreensão da proposta de automatização da detecção de fraude na emissão de certificados digitais são apresentados os conceitos: Infraestrutura de Chaves Públicas 2.1, Classificação 2.2 e Clusterização 2.3.

2.1. Infraestrutura de Chaves Públicas

Uma Infraestrutura de Chaves Públicas (ICP) é um órgão que tem como objetivo manter uma estrutura de emissão de chaves públicas. No caso da ICP-Brasil, Infraestrutura de Chaves Públicas Brasileira, esta aplica políticas de certificação, protocolos técnicos, regimes normativos, procedimentos, trabalhando com a relação de confiança com as partes que utilizem certificados digitais. Uma ICP desempenha também a tarefa de gerenciar o ciclo de vida dos certificados, uma vez que certificados podem ser revogados, como no caso de comprometimento da chave privada de determinado titular de um certificado digital. A primeira função de uma ICP é permitir, por meio das Autoridades Certificadoras, a distribuição e o uso de chaves públicas e certificados com garantia de segurança.

A ICP-Brasil foi criada com base no modelo de ICP hierárquica, e foi instituída de acordo com a Medida provisória 2.200-2 de 24 de agosto 29 de 2001, para garantir autenticidade, integridade e validade jurídica de documentos eletrônicos. É composta por: um Comitê Gestor, uma AC raiz, ACs e ARs. O Comitê Gestor tem autoridade para gerir políticas a serem executadas pela AC Raiz, papel realizado pelo Instituto Nacional de Tecnologia da Informação (ITI) [ITI 2020a].

A AC Raiz, nesse caso o ITI, tem a maior autoridade e está no topo da hierarquia de certificação. Cabe a essa, executar as políticas, normas técnicas e operacionais impostas pelo comitê gestor. Segundo a Medida provisória 2.200-2 a AC Raiz tem a função de emitir, distribuir, revogar e gerenciar certificados das ACs imediatamente subsequentes, assim como executar atividades de fiscalização e auditoria. As ACs são entidades autorizadas a emitir certificados digitais e vincular determinado titular ao par de chaves criptográficas. Essas têm responsabilidade por emitir, expedir, distribuir, revogar e gerenciar os certificados; além de disponibilizar aos usuários listas de certificados revogados, assim como outras informações válidas, mantendo registro de suas operações. As ARs são entidades vinculadas a determinadas ACs. Cabe às ARs identificar e cadastrar usuários que devem comparecer presencialmente, além de encaminhar solicitações de certificados às ACs e manter registros de suas operações [ITI 2020a].

2.2. Classificação

Classificação é uma técnica de aprendizado de máquina supervisionada utilizada para a criação de um modelo com base na análise de um conjunto de dados de treinamento. Através de registros já rotulados, o algoritmo procura um padrão para classificação de novos elementos. O modelo criado é utilizado para predição do valor da classe, também chamado de rótulo/label/classe, em novos registros. Existem diferentes técnicas para criação de um modelo, utilizando, por exemplo, regras de classificação, árvores de decisão, fórmulas matemáticas, redes neurais, classificação bayesiana, máquinas de vetores de suporte e k-NN [Han et al. 2011].

A classificação requer um conjunto de dados que já possua as classes de seus registros. Esse conjunto de dados será dividido em treinamento e teste. A parte de treinamento é usada para criar um modelo do conjunto de dados e a versão de teste é usada para validar esse modelo. É necessário dividir o conjunto de dados várias vezes aleatoriamente e, em seguida, calcular a média dos resultados, pois se feito apenas uma vez, as partes de treinamento/teste poderão não representar todos os dados.

2.3. Clusterização ou Agrupamento

A clusterização ou agrupamento é um método de aprendizado de máquina não supervisionado. Diferente da classificação, os dados de entrada não possuem uma classe que os rotula. Ele divide os dados em grupos, denominados clusters, onde cada cluster possui os pontos que são mais similares entre si do que aos de outros grupos. Os dois algoritmos de clusterização mais conhecidos são o DBSCAN e o k-means.

O DBSCAN é um algoritmo de clusterização baseado em densidade, onde a quantidade de clusters não é fixa antes da execução [Ester et al. 1996]. Normalmente é utilizado para agrupar pontos pertencentes a um plano cartesiano. Os pontos de entrada do algoritmo podem ser classificados em três diferentes tipos: centrais, de borda e ruídos.

Os pontos centrais que possuem um número de vizinhos maior ou igual a MIN_POINTS. Os pontos de borda que tem um número de vizinhos menor que MIN_POINTS, mas é alcançável por algum ponto central. E os ruídos que apresentam um número de vizinhos menor que MIN_POINTS e não é alcançável por nenhum ponto central.

Sua execução depende de dois atributos, o EPS, que é a distância de máxima entre dois pontos para serem considerados vizinhos, e o MIN_POINTS, que é a quantidade mínima de vizinhos que um ponto deve ter para ser classificado como central. O algoritmo funciona percorrendo os pontos e calculando seus vizinhos. Para os pontos centrais, o algoritmo busca os pontos de borda vizinhos. Um cluster pode possuir mais de um ponto central desde que estes sejam vizinhos. Ao final, os pontos inalcançáveis são considerados ruídos. Para este trabalho, foi utilizado um algoritmo DBSCAN adaptado para comparação entre objetos ao invés de pontos em um plano cartesiano. Para comparação de distância, foi desenvolvida uma métrica que considera os atributos dos dossiês.

3. Trabalhos Correlatos

Existem diversos trabalhos na literatura propondo métodos de identificação de fraudes de identidade. A maioria das fraudes de identidade está relacionada ao uso ilegal de contas financeiras ou de cartões de crédito. A maior parte dos trabalhos presentes no estado da arte propõe métodos de detecção de fraude de cartão de crédito (ex. [Thennakoon et al. 2019, Awoyemi et al. 2017]), de telecomunicações (ex. [Zhao et al. 2018]) e de intrusão por computador (ex. [da Costa et al. 2019]), como mostrado em [Kou et al. 2004, Phua et al. 2010a, Padhi et al. 2020]. Em [Phua et al. 2010a], eles categorizam os trabalhos que utilizam aprendizado de máquina em dados rotulados ou não rotulados e na utilização de algoritmos supervisionados, não supervisionados, semi-supervisionados ou mais de um algoritmo em uma proposta híbrida. Em [Bolton and Hand 2002], eles também apresentam uma proposta com ferramentas estatísticas para detecção de fraude utilizando aprendizado de máquina.

Além disso, esses trabalhos podem ser subdivididos em dois tipos: fraudes de transação/comportamental e fraudes de aplicação. A fraude de aplicação ocorre quando alguém tenta emitir um certificado de identidade (ex. passaporte, cartão de crédito) utilizando a identidade de outra pessoa. A fraude na transação ocorre quando um ladrão de identidade realiza algumas operações ou transações usando uma identidade falsa, ou roubada. No entanto, a maioria dos trabalhos na área de fraude foram propostos para a detecção de fraudes transacionais. O estado da arte da detecção de fraudes de aplicação é limitado, apenas alguns trabalhos foram propostos [Kshirsagar and Dole 2014].

Wheeler et al. em [Wheeler and Aitken 2000] propõe o uso de raciocínio baseado em casos para a detecção de fraudes na emissão de cartão de crédito. Como [Wheeler and Aitken 2000], Phua et al. em [Phua et al. 2009] propõem um método para detectar fraudes de aplicações de crédito com base na similaridade entre aplicações novas e anteriores usando uma matriz de ponderação. Sua técnica proposta difere-se principalmente ao considerar diferenças temporais e espaciais na medida de similaridade, chamada "Communal Analysis Suspicion Scoring" (CASS). Como é simulado para execução em tempo real, o CASS não considera os rótulos das classes ao pontuar aplicações. Ele usa apenas rótulos para determinar a eficácia de sua abordagem.

A literatura sobre clusterização/detecção de anomalias é bastante vasta, con-

tendo vários métodos, como, por exemplo, [Lee and Stolfo 1998, Lee et al. 2000, Chandola et al. 2009]. O objetivo de tais métodos é separar padrões normais e anormais [Agrawal and Agrawal 2015]. Em [Abdelhalim and Traoré 2010], eles propõem uma estrutura não supervisionada para detecção de fraude de identidade. O detector de fraude é composto por uma árvore de decisão que classifica a aplicação em: muito suspeita, normal, pouco suspeita ou fraudulenta. O os padrões relacionados extraídos e o conhecimento especializado sobre fraude são usados para criar regras para compor a árvore. Seguindo a mesma ideia de evitar fraudes em aplicações, em [Phua et al. 2010b] para combater esse problema, eles propõem um sistema de detecção de fraudes com multiplas camadas. Este sistema usa dois algoritmos para detectar fraudes: Detecção Comum (CD) e Detecção de Spike (SD). O CD verifica a ligação entre de múltiplos atributos e o SD verifica a ligação de um único atributo. Esses dois métodos geram uma pontuação, quanto maior a pontuação de uma aplicação, maior a probabilidade de ser uma fraude. Existem outros trabalhos como [Kshirsagar and Dole 2014, Dutta et al. 2017] que seguem a mesma ideia e usam CD e SD para detectar fraudes em aplicações. No entanto, nenhum dos trabalhos apresentados na literatura propôs uma estrutura para detecção de fraudes de aplicação para emissão de certificados digitais. Na próxima seção, é apresentada a situação atual da emissão de certificados digital no Brasil e em seguida a proposta deste trabalho para evitar fraudes de identidade na aplicação para emissão de um certificado digital.

4. Cenário Atual

A etapa inicial do processo de emissão é a validação presencial, na qual o requerente vai pessoalmente a uma unidade da AR para validar os dados preenchidos na solicitação do certificado digital. Para isso é feito o agendamento diretamente com a AR que informará ao solicitante quais são os documentos necessários [ITI 2020a]. Um agente de registro (AGR) fica responsável pela execução desta verificação e validação. Essa etapa é crucial para a prevenção de fraudes, é nesse momento que o fraudador vai pessoalmente até o ponto de atendimento. Para evitar que fraudes aconteçam, o ITI criou um documento com os procedimentos que devem ser executados na verificação presencial.

4.1. Validação presencial

Os procedimentos para identificação do requerente e comunicação de irregularidades no processo de emissão de um certificado digital ICP-Brasil pode ser consultado em [ITI 2020b]. Nesse documento são descritas as atividades que devem ser executadas no processo de Validação Presencial. Em suma, como apresentado na Figura 1, primeiramente o AGR faz a verificação dos documentos de identidade para comprovar que o indivíduo é realmente aquele cujos dados constam na documentação apresentada, tanto para certificados de pessoa física quanto jurídica (representante legal da organização). Os documentos solicitados são: cédula de identidade com foto (normalmente RG ou CNH), Passaporte, Carteira Nacional de Estrangeiro (CNE), Comprovante de residência/domicílio, emitido há no máximo três meses da data da validação presencial, e caso necessário, PIS/NIS/PASEP/NIT ou carteiras de conselhos profissionais. Os AGRs devem fazer uma análise minuciosa da cédula de identificação, principalmente do RG e CNH, incluindo consultas às bases oficiais, auxílio de peritos e/ou softwares para validação da veracidade do documento apresentado. Caso a AR conclua pela validade do documento de

identificação, deve prosseguir com o processo de emissão do certificado digital. Caso a AR conclua pela não validade do documento, deve comunicar a AC para que essa faça o comunicado de tentativa de fraude ao ITI. O AGR também coleta as impressões digitais e fotografia do requerente.

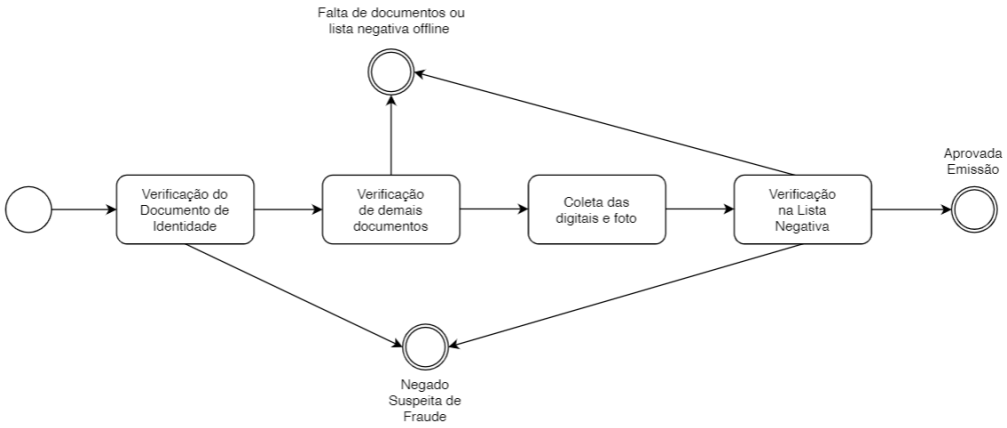

Figura 1. Funções do AGR no processo de validação presencial

As ACs disponibilizam, para suas ARs vinculadas, uma interface para consulta a base de dados da Lista Negativa da AC.

> "Lista Negativa é um conjunto de informações derivadas dos comunicados de fraude, ou indícios de fraude, feitos pelas AC (ou pelo próprio ITI através de auditoria e fiscalização) da ICP-Brasil ao ITI, em que contém o modo de operação da ocorrência, as informações biográficas do documento apresentado e, se for o caso, das informações sobre a empresa, características fisiológicas do suposto fraudador, a imagem da face e do documento de identificação utilizado pelo suposto fraudador" [ITI 2020a].

Caso ocorra qualquer indisponibilidade no banco de dados da Lista Negativa da AC, o certificado digital não deve ser emitido. A interface para consulta à Lista Negativa disponibilizada para os AGR apresenta algumas consultas pré-definidas, tais como, consulta aos dez maiores supostos fraudadores da ICP-Brasil, aos comunicados de indícios ou fraudes dos últimos sete dias. Essas consultas, consideram a unidade da federação (UF) e cidade em que ocorreu o indício ou fraude, relato e data da ocorrência, modo como foi detectado o indício ou fraude, dados apresentados no documento de identificação da pessoa física, características físicas (cor dos olhos, pele, cor do cabelo, deficiências aparentes, tatuagens, marcas, cicatrizes, idade aparente, gênero, tipo do cabelo), informações da empresa, imagem de todo documento de identificação da ocorrência. Com essas informações dos fraudadores, o AGR pode procurar na base por alguém similar à pessoa que está tentando emitir o certificado. Caso a pesquisa apresente muitos resultados, e não haja certeza sobre a inclusão de outras características físicas, os AGRs devem relacionar essa pesquisa a outros campos.

Caso não se obtenha qualquer resultado, deve ser realizada uma busca por fraudadores na região em que a AR está operando. Essa região pode, também, estender-se por UFs próximas ou mais específicas como a municípios próximos. Se os resultados das

pesquisas concluírem pela ausência do requerente do certificado digital na Lista Negativa, os AGRs devem prosseguir com as validações, verificações e finalização do dossiê. Na situação em que os resultados das consultas constatem que o requerente do certificado digital integra a Lista Negativa, com a imagem da face e/ou do documento de identificação coincidente com o apresentado pelo requerente, os AGRs devem realizar as validações e verificações elencadas, preferencialmente, comunicar à AC vinculada para que se faça uma análise detalhada do caso. Caso a AR e/ou a AC conclua pela não emissão do certificado digital, a AC deve comunicar a tentativa de fraude ao ITI.

4.2. Limitações e Falhas no Processo

Como pode ser visto na Figura 1, o AGR é responsável por diversas tarefas de extrema importância para detecção da fraude, sendo assim, a principal figura deste cenário. Quando um AGR é um funcionário corrompido, nada o impede de manipular as etapas do processo, enviando dados que não condizem com os do requerente. Além disso, por mais que o AGR não tenha a intenção de corromper o sistema, qualquer desatenção ou displicência nas etapas de verificação, podem indiretamente colaborar com a tentativa de fraude. Pode ser concluído, observando o cenário atual, que muitas responsabilidades e confiança são dadas para apenas uma pessoa, pessoa essa que pode não ser honesta e está, como qualquer ser humano, sujeita a falhas. Para corroborar com essa questão, foram analisados dados extraídos de dossiês reais, na busca de possíveis inconsistências. Esses dados foram fornecidos, com todas os cuidados necessários, por um ente da ICP-Brasil de maneira anonimizada, sem colocar em risco a privacidade de nenhum indivíduo. Foram enviadas apenas características físicas presentes em 36.704 dossiês. Após análise, encontramos algumas inconsistências como: cor do cabelo preenchida com diversas opções (exemplo, escuro, loiro, ruivo, grisalho, branco), a cor dos olhos preenchidas com claro e escuro ao mesmo tempo, idade aparentemente preenchida com duas faixas etárias controversas (exemplo, entre 30 e 50 anos e mais de 50 anos ou menor que 30 anos e entre 30 e 50 anos) e gênero selecionado com a opção feminino e masculino. Essas informações passadas erroneamente prejudicam muito o sistema atual, visto que o mesmo trabalha com essas informações para detectar fraudes.

Atualmente, as informações coletadas ficam apenas no dossiê. Esse dossiê é composto por arquivos escaneados. Para coletar os dados, foi necessário a utilização de uma tecnologia de reconhecimento de caracteres em imagens, também chamada de OCR. Devido à qualidade dos arquivos escaneados, muitos dados não foram capturados por completo. Visto isso, para que a qualidade dos dados seja mantida, é essencial que eles sejam armazenados quando coletados em uma base de dados. Um ponto de falha no processo que foi considerado grave, está na captura das características físicas. As pessoas mudam a cor de cabelo, podem utilizar lentes para mudar a cor dos olhos, esconder tatuagens ou cicatrizes. Para contornar esses problemas, este trabalho propõe a extração dessas características através de algoritmos de reconhecimento facial que consideram também as distâncias e proporções faciais. Um exemplo de ferramenta que faz esse trabalho é o Amazon Rekognition que analisa uma imagem e compará-la a outras imagens para verificar se são a mesma pessoa [Amazon 2020].

5. Proposta

A proposta deste trabalho consiste na utilização de algoritmos de aprendizado de máquina para automatização da detecção de fraude na emissão de certificados digitais. Como pode

ser visto na Figura 2, é proposta a utilização de uma abordagem híbrida fazendo uma junção de técnicas de clusterização e classificação. O trabalho optou pela escolha de uma abordagem híbrida pelo fato de que a clusterização identifica ruídos e utiliza a similaridade para agrupar os dados e a classificação cria modelos preditivos. Com a similaridade, conseguimos identificar padrões e assim criar os rótulos não binários, que representam grupos com características distintas e não somente grupos de fraudadores e não fraudadores. Os ruídos ajudam a identificar padrões ainda não identificados nos fraudadores já detectados. A utilização da classificação permite a criação de modelos que rotulem novas aplicações em tempo real. Devido a grande quantidade de métodos de classificação, é possível ter mais de uma opinião sobre a classe do requerente.

Quando o requerente chega presencialmente no ponto de atendimento da autoridade registradora, após a verificação dos documentos, e coleta das digitais e foto, o AGR envia a foto para uma análise de reconhecimento facial no banco de dados dos fraudadores. O objetivo é encontrar alguma face similar a do requerente, como pode ser visto em 1 na Figura 2. Caso alguma foto seja encontrada, uma notificação de tentativa de fraude é enviada para análise, caso contrário, o processo segue.

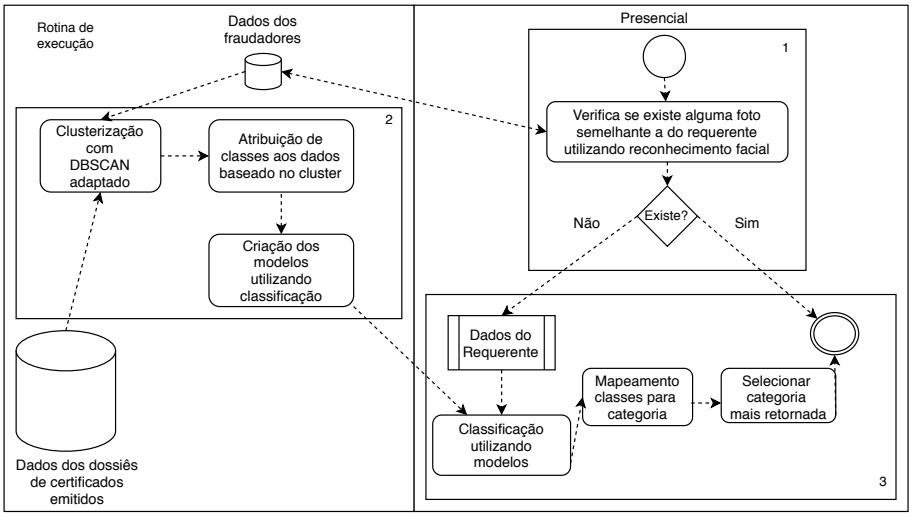

Figura 2. Desenho da proposta

Os dados de treinamento utilizados para construir o modelo são compostos por informações extraídas dos dossiês até um certo período, definido pela autarquia federal, junto a dados dos fraudadores. A proposta é rodar diariamente o algoritmo de clusterização, visto que novos fraudadores são detectados e os modus operandi mudam e se adaptam com o tempo. Como visto em 2 na Figura 2, para rotular os dados utilizando como parâmetro o nível de periculosidade do requerente, foi utilizado o algoritmo de clusterização DBSCAN com a métrica de similaridade adaptada para comparar os atributos ao invés da utilização somente da distância Euclidiana, como de costume. Visto que as características físicas já foram verificadas, dentre os dados presentes no dossiê, foram selecionados os atributos AR, AGR, unidade da AR, distância entre o CEP do requerente e o da unidade da AR, gênero, idade, CNAE e natureza jurídica, os dois últimos no caso de certificados para pessoa jurídica. A métrica compara a maioria dos atributos de maneira binária, somando 0 caso sejam iguais e 1 caso sejam diferentes, menos a idade

e a distância entre a localização da residência do requerente e a localização da unidade da AR que foram calculadas utilizando a diferença e a distância Euclidiana, respectivamente. O valor da distância foi normalizado para que os valores ficassem entre 0 e 1. Essa medida de similaridade determina quanto um dossiê está perto de outro, é mais similar, quanto mais próximo de zero, maior a chance de estarem no mesmo cluster. Outros dados podem ser adicionados a essa medida, foram selecionados estes para este trabalho, pois, foram mais fáceis de extrair dos documentos sem apresentar erros. No entanto, também é interessante capturar informações acerca do certificado, como, por exemplo, seu tipo e validade.

Após rodar o algoritmo de clusterização, o cluster no qual o dossiê pertence se torna o rótulo para a posterior criação dos modelos utilizando diferentes algoritmos de classificação. Com os modelos criados, aqueles algoritmos que tiverem a melhor acurácia e tempo de execução são utilizados na etapa presencial para classificar o requerente, como visto em 3 na Figura 2. Após classificado é feito um mapeamento da classe, que no momento é cluster, para uma categoria correspondente. Quanto maior a quantidade de fraudadores no cluster maior a chance de que dossiês similares a ele sejam fraudes também. Cada cluster é categorizado dependendo da quantidade de fraudadores que foram agrupados nele. Essa categoria é usada para rotular o registro do requerente em relação a sua periculosidade. As categorias são: risco alto, risco médio, risco moderado e risco baixo. Na categoria alto risco estão os ruídos (anomalias) e os clusters com uma porcentagem mais alta. A categoria risco médio apresenta uma quantidade ainda significante de fraudadores. A categoria moderada apresenta uma quantidade de dados de fraudadores que não é considerada muito perigosa. E finalmente, risco baixo são aqueles cluster que a quantidade de fraudadores é muito pequena ou nula. A quantidade de fraudadores por cluster necessária para categorizá-lo é um parâmetro que deve ser estipulado pela autarquia federal. Vale ressaltar que uma baixa tolerância de porcentagem fraudes por cluster pode criar falsos negativos e uma tolerância muito alta pode não identificar possíveis fraudes. Uma fraude detectada pode evitar prejuízos muito grandes. É melhor gerar um falso negativo do que não detectar um falso positivo. Como são utilizados mais de um modelo para classificar o requerente, a categoria, que representa a classe/cluster, que mais aparece é utilizada como categoria final. Em caso de empate, a categoria do modelo que foi gerado pelo algoritmo com melhor desempenho é escolhida.

6. Resultados

A maneira mais comum de avaliar estruturas como a proposta neste artigo consiste em usar base de dados públicas contendo informações reais sobre fraude de identidade. No entanto, não existem dados públicos de dossiês, visto que esses dados podem colocar em risco a privacidade de seus donos. Uma abordagem alternativa seria a geração de dados sintéticos de dossiês, porém não seria possível verificar a eficácia da proposta. Para avaliar o método de automatização de detecção de fraude na emissão de certificados digitais, foi contado com a ajuda de uma autoridade certificadora vinculada a ICP-Brasil. Os experimentos foram rodados dentro da empresa, tendo apenas os resultados finais enviados. Com isso, a identidade de todos os indivíduos permaneceram seguras, respeitando a confidencialidade e privacidade dos clientes. O ambiente de onde os experimentos foram realizados é composto por um processador Intel Core I7 e 32GB de RAM. Os códigos foram escritos em Python utilizando as bibliotecas pandas e scikit-learn (para os algoritmos

de classificação). Essa AC disponibilizou dossiês do ano de 2017 para os testes. Foram recuperados um total 36.704 dossiês de certificados de pessoa jurídica, todos atendendo os critérios de auditoria do ITI. Visto que os dossiês são compostos por arquivos escaneados, alguns não foram capturados devido a qualidade da imagem estar abaixo do necessário para reconhecimento através do algoritmo de OCR. Não foi possível ter acesso aos dados dos fraudadores, sendo assim, foi criado um cenário de fraude onde um existia um AGR corrompido e criado 370 fraudes sintéticas relacionadas a esse AGR para integrar o montante de 36.704 dossiês totalizando em 37.074 registros. Os atributos utilizados foram: AR, AGR, unidade da AR, distância entre o CEP do requerente e o da unidade da AR, gênero, idade, CNAE e natureza jurídica.

Para execução do algoritmo DBSCAN, é necessário a utilização de dois parâmetros. O primeiro, chamado EPS, referente à distância máxima entre dois pontos para serem considerados vizinhos. O segundo, a quantidade mínima de pontos de borda vizinhos para um ponto ser considerado central. Para evitar que os parâmetros fossem definidos de maneira arbitrária, o algoritmo foi executado com configurações diferentes para 10% dos registros (3670). A execução foi realizada variando os parâmetros de EPS entre 0 e 7 em uma combinação com a variação da quantidade mínima de pontos entre 10 e 100. O objetivo foi encontrar uma configuração que maximizasse o número de clusters e minimizasse o ruído, como pode ser visto no gráfico da Figura 3 no ponto mais alto da linha azul e no ponto mais baixo da linha vermelha. Os valores escolhidos para o EPS resultam da distância mínima entre dois dossiês, que seria zero, e a distância máxima, que dentro da medida de similaridade utilizada, acontecia quando todos os sete atributos utilizados eram diferentes.

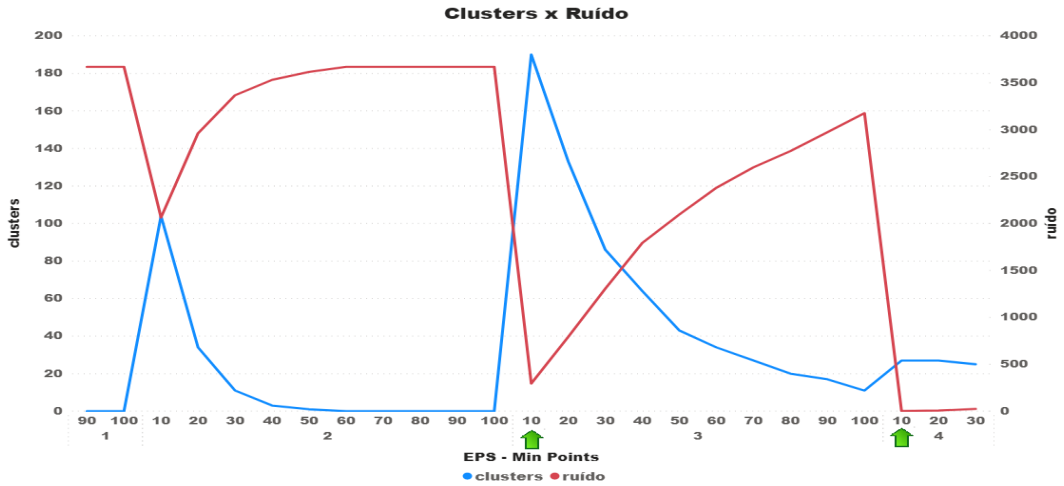

Figura 3. Quantidade de cluster e ruídos por valor de min_points

Na execução dos testes, foi possível verificar duas configurações com maximização de clusters e minimização de ruídos. Na primeira, com EPS igual a 3 e mínimo de pontos igual a 10, foram encontrados 190 clusters e 295 dossiês como ruído. Para o EPS igual a 4 e mantendo a quantidade de pontos mínimos, foram encontrados 27 clusters e 3 dossiês como ruído. Uma vez que com o aumento na quantidade da entrada a tendência é uma diminuição na porcentagem de ruídos, a primeira configuração foi a escolhida.

Após executar o DBSCAN para todos os registros, os dossiês foram divididos em

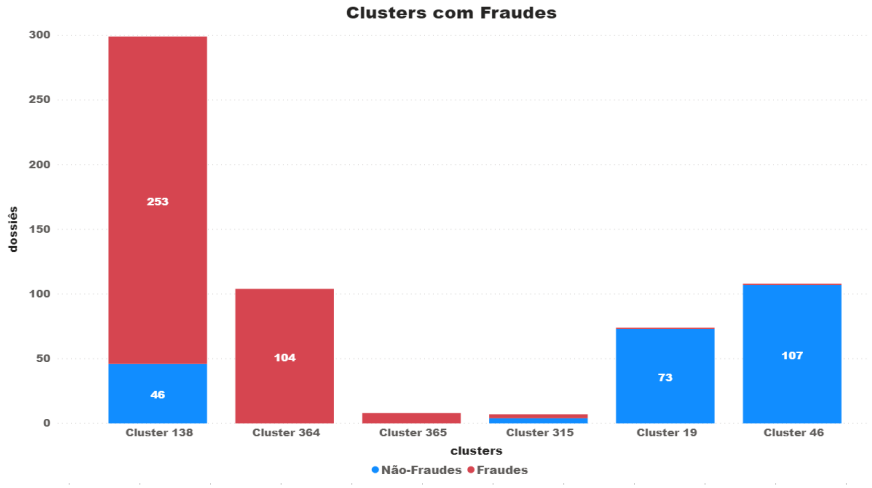

Figura 4. Cluster com fraudes

365 clusters, tendo encontrado apenas 24 como ruídos. Como pode ser visto na Figura 4, dos clusters encontrados, seis deles tiveram pelo menos uma das fraudes sintéticas. Vale destacar os resultados encontrados em dois destes clusters, onde a porcentagem de fraude encontrada foi de 84.15% e 100%, respectivamente, existindo mais de 100 registros em cada um. Devido às altas porcentagens, e quantidade de registros dos clusters, é provável que estes dois fossem classificados como clusters de alta periculosidade. Outros dois clusters com porcentagens expressivas, com 100% e 42.86% de fraudes, poderiam receber a classificação de alta periculosidade também, mas ambos com menos de 10 dossiês em casa cluster.

	Algoritmo	Execução	Acurácia
1	One Class SVM	18.42 s	2%
2	Logistic Regression	53.08 s	5%
3	Gaussian NB	0.15 s	12%
4	Linear Discriminant Analysis	0.09 s	20%
5	Extra Tree Classifier	0.15 s	62%
6	K-Neighbors Classifier	0.31 s	68%
7	Decision Tree Classifier	0.29 s	76%
8	Extra Trees Classifier	14.82 s	79%
9	Bagging Classifier	2.31 s	79%
10	Random Forest Classifier	13.60 s	81% %

Tabela 1. Resultado dos algoritmos de classificação.

O objetivo de testar diversos algoritmos de classificação era encontrar aqueles que tivessem boas métricas e um tempo rápido de execução. O tempo de execução está considerando o treinamento, construção do modelo e teste. Os dados foram divididos em 70% para treinamento e 30% para teste. Como pode ser visto na Tabela 1, os algoritmos, enumerados de 5 até 10, apresentaram o melhor resultado com relação a acurácia e o tempo de execução, que não ultrapassou 14.82 segundos para o mais demorado deles. Visto que todos esses algoritmos são rodados para construir modelos de predição, na prática, levaria para cerca de 37 mil registros, se rodados paralelamente 14.82 segundos para que todos fossem construídos. Temos hoje em dia 3,5 milhões de certificados emitidos no Brasil

por ano, para rodar um ano demoraria 23 minutos para construir todos os modelos com os algoritmos que tiveram melhor desempenho. Com os modelos prontos, a avaliação da periculosidade do requerente é quase instantânea. Eles podem ser enviados diariamente para as Autoridades Certificadoras, para que sempre estejam atualizados com os dossiês novos de certificados emitidos e as novas tentativas de fraude detectadas.

7. Conclusão

Questões de fraude de identidade vêm sido abordadas em trabalhos na literatura a anos. Poucos trabalhos apresentam soluções para problemas de aplicação, onde o fraudador tenta emitir um certificado de identidade em nome de outra pessoa. Nenhum trabalho na literatura havia abordado a questão de fraude na emissão de certificados digitais. Ao analisar o processo, foram encontrados alguns problemas e a maioria deles relacionados a confiança atribuída a uma única pessoa, o AGR. O processo foi analisado como lento, custoso, e suscetível a falha. Com isso, foi sugerida uma automatização do processo de detecção de fraude utilizando aprendizado de máquina em uma abordagem híbrida. O método tem como objetivo categorizar a aplicação do requerente ao nível de periculosidade, ou seja, nas chances de ser uma fraude. O método proposto foi testado com dados de dossiês reais. Os dados dos fraudadores eram sintéticos criados em um cenário onde existia um AGR corrompido. Os resultados foram satisfatórios mostrando que é possível reduzir o tempo de execução, diminuindo assim, custos e aumentando a segurança do processo. Como trabalhos futuros, é sugerido o teste dessa proposta com dados de fraudadores reais e adição de mais atributos na medida de similaridade.

8. Acknowledgments

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) - Código de Financiamento 001 - e também financiado pela Newton Advanced Fellowship - Royal Society.

Referências

Abdelhalim, A. and Traoré, I. (2010). Unsupervised identity application fraud detection using rule-based decision tree. In *SEDE*, pages 261–268.

Agrawal, S. and Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 60:708–713.

Amazon (2020). Detecte, analise e compare rostos.

Awoyemi, J. O., Adetunmbi, A. O., and Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 International Conference on Computing Networking and Informatics (ICCNI)*, pages 1–9. IEEE.

Bolton, R. J. and Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, pages 235–249.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., and de Albuquerque, V. H. C. (2019). Internet of things: A survey on machine learning-based intrusion detection approaches. *Computer Networks*, 151:147–157.

Dutta, S., Gupta, A. K., and Narayan, N. (2017). Identity crime detection using data mining. In *2017 3rd International Conference on Computational Intelligence and Networks (CINE)*, pages 1–5. IEEE.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.

Han, J., Kamber, M., and Pei, J. (2011). Data mining concepts and techniques third edition. *Morgan Kaufmann*.

ITI (2020a). Instituto Nacional de Tecnologia da Informação.

ITI (2020b). Procedimento de Identificação do Requerente.

Kou, Y., Lu, C.-T., Sirwongwattana, S., and Huang, Y.-P. (2004). Survey of fraud detection techniques. In *IEEE International Conference on Networking, Sensing and Control, 2004*, volume 2, pages 749–754. IEEE.

Kshirsagar, A. and Dole, L. (2014). Recognizing the theft of identity using data mining. *International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001: 2008 Certified Journal, Volume 4, Issue 4*.

Lee, W. and Stolfo, S. (1998). Data mining approaches for intrusion detection.

Lee, W., Stolfo, S. J., and Mok, K. W. (2000). Adaptive intrusion detection: A data mining approach. *Artificial Intelligence Review*, 14(6):533–567.

Padhi, B., Chakravarty, S., and Biswal, B. (2020). Anonymized credit card transaction using machine learning techniques. In *Advances in Intelligent Computing and Communication*, pages 413–423. Springer.

Phua, C., Gayler, R., Lee, V., and Smith-Miles, K. (2009). On the communal analysis suspicion scoring for identity crime in streaming credit applications. *European Journal of Operational Research*, 195(2):595–612.

Phua, C., Lee, V., Smith, K., and Gayler, R. (2010a). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Phua, C., Smith-Miles, K., Lee, V., and Gayler, R. (2010b). Resilient identity crime detection. *IEEE transactions on knowledge and data engineering*, 24(3):533–546.

Sproule, S. and Archer, N. (2007). Defining identity theft. In *Eighth World Congress on the Management of eBusiness (WCMeB 2007)*, pages 20–20. IEEE.

Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., and Kuruwitaarachchi, N. (2019). Real-time credit card fraud detection using machine learning. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 488–493. IEEE.

Wheeler, R. and Aitken, S. (2000). Multiple algorithms for fraud detection. In *Applications and Innovations in Intelligent Systems VII*, pages 219–231. Springer.

Zhao, Q., Chen, K., Li, T., Yang, Y., and Wang, X. (2018). Detecting telecommunication fraud by understanding the contents of a call. *Cybersecurity*, 1(1):8.