Dinâmica da digitação aplicada à autenticação periódica de usuários em ambientes virtuais de aprendizagem: um estudo de caso com o Moodle

Marco A S Cruz*, Otavio A M Camargo, Julio Cesar Duarte, Ronaldo Goldschmidt
Instituto Militar de Engenharia, Praça General Tibúrcio, 80, 22290-270
Praia Vermelha, Rio de Janeiro, RJ, Brasil
***marco.aurelio.s.cruz@gmail.com**

RESUMO: A autenticação de usuários em Ambientes Virtuais de Aprendizagem (AVAs) geralmente exige uma senha para conexão. Tal método é incapaz de garantir a autenticidade de cada atividade do usuário após a autenticação. Para mitigar esse problema, Cruz et al. propôs um mecanismo de autenticação periódica de usuários em AVAs [1]. Foram aplicadas técnicas de aprendizado de máquina para construir modelos de reconhecimento baseados na dinâmica da digitação dos usuários. Para demonstrar sua viabilidade em um cenário real, com um grande número de usuários, este trabalho conduziu um estudo de caso aplicado, usando o Moodle, a um grupo de 307 usuários, produzindo um total de 4.829 palavras avaliadas. Cerca de 89% dos modelos de reconhecimento atingiram pelo menos 80% de acurácia, indicando a eficácia do mecanismo na identificação de autorias suspeitas.

PALAVRAS-CHAVE: Aprendizado de Máquina. Reconhecimento de Usuários. Dinâmica da Digitação. Ambiente Virtual de Aprendizagem. Moodle.

ABSTRACT: The authentication of users in Virtual Learning Environments (AVAs) usually requires a password to connect to the environment. Such method is unable to ensure the authenticity of every user activity. In order to mitigate this problem, Cruz et al. proposed an engine to execute periodic and non-intrusive authentication of users in VLE [1]. Machine learning techniques were applied to build recognition models based on the keystroke dynamics of users and it is VLE independent. In order to demonstrate its practical feasibility in a real scenario with a large number of users, this paper conducted an applied case study, using Moodle, to a group of 307 users, producing a total of 4,829 evaluated strings. About 89% of the recognition models achieved at least 80% of accuracy, indicating the effectiveness of the engine at identifying suspicious authorships.

KEYWORDS: Machine Learning. User Recognition. Keystroke Dynamics. Virtual Learning Environments. Moodle.

1.Introdução

A Educação a Distância (EaD) é uma modalidade de ensino que vem crescendo significativamente nos últimos anos em todo o mundo [2]. O censo sobre a Educação superior brasileira, realizado em 2015 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), revela que o número de matrículas em cursos de graduação à distância de instituições públicas e privadas cresceu, respectivamente, 16,3% e 41,2%, entre 2013 e 2014 [3].

A evolução das tecnologias da informação e comunicação (TIC) em muito tem contribuído para apoiar o amadurecimento da EaD [4]. Entre tais TICs, estão os chamados Ambientes Virtuais de Aprendizagem (AVAs). Segundo [5], AVAs são "tecnologias integradoras e abrangentes capazes de organizar e definir um ponto focal para os processos educacionais mediados por computador, apoiar os professores e promover o engajamento dos alunos". Em geral, os AVAs oferecem recursos síncronos e assíncronos para que estudantes e tutores possam interagir, independentemente da disponibilidade de tempo e da localização física de cada um.

Atualmente, existem diversos AVAs que podem ser adotados sem custo pelas instituições de ensino [6]"mendeley":{"formattedCitation":"[6]","plainTextFormattedCitation":"[6]","previouslyFormattedCitation":"[6]"},"properties":{"noteIndex":0},"schema":"https://github.com/citation-style-language/schema/raw/

master/csl-citation.json"}. Entre os mais conhecidos no Brasil estão o *Moodle*, o *Teleduc* e o *AulaNet* [5]. Parte deles se mantém em constante atualização, procurando suprir demandas de seus usuários nos mais variados aspectos tais como portabilidade, usabilidade, segurança, dentre outros [7].

Apesar de todos os avanços proporcionados pelos AVAs na EaD, uma das principais questões acerca desta modalidade de ensino é quanto à autenticidade dos usuários dessas plataformas [8]–[10]. Em geral, a autenticação de um usuário junto a um AVA é intrusiva e pontual, ocorrendo no momento em que o usuário se conecta ao ambiente, mediante a digitação de uma senha [10]–[12]. No entanto, essa abordagem permite que usuários não credenciados, após a autenticação inicial, assumam o papel de usuários credenciados, o que pode ocasionar diversos problemas, como, por exemplo: falhas de segurança (por meio de acesso não autorizado a determinados conjuntos de informações e de usuários) ou distorções sobre a percepção do desempenho acadêmico dos estudantes (ao permitir que tarefas e atividades propostas no ambiente sejam desenvolvidas por usuários diferentes daqueles que, de fato, deveriam desenvolvê-las) [8], [9], [12], [13].

Com o objetivo de aumentar a segurança e mitigar os problemas mencionados nos AVAs, *Cruz et al.* [1] propôs um mecanismo que permite verificações de autoria periódicas e não intrusivas em AVAs.

O mecanismo citado usa técnicas de aprendizado de máquina para construir modelos de reconhecimento baseados na dinâmica de pressionamento de tecla dos usuários e é independente do AVA. Embora este tenha apresentado resultados preliminares promissores foi avaliado em um contexto restrito a 17 usuários.

Portanto, a fim de demonstrar sua viabilidade prática em um cenário real com mais usuários e dados, este artigo relata um estudo de caso no qual o mecanismo foi integrado ao Moodle e aplicado a um grupo de 307 usuários. Mais de 1,6 milhão de toques foram coletados.

O artigo apresenta mais seis seções. A seção 2 apresenta os conceitos básicos sobre dinâmica de teclas e aprendizado de máquina, necessários para entender o mecanismo de autenticação periódica do usuário usado em nosso estudo de caso. A seção 3 discute trabalhos relacionados, comparando-os com o motor proposto por [1] e com o estudo de caso relatado no presente trabalho. As principais características desse mecanismo estão resumidas na seção 4. A seção 5 contém os detalhes do protótipo implementado. Detalhes sobre o estudo de caso e seus resultados são apresentados na seção 6. A seção 7 conclui o trabalho, resumindo as principais conclusões e sugerindo possibilidades para trabalhos futuros.

2 Fundamentos

2.1 Reconhecimento de usuários baseado na dinâmica da digitação

Existem basicamente três formas de se autenticar usuários em sistemas de informação [14]:

- a partir de recursos de que o usuário disponha (e.g. a utilização de *tokens*);
- a partir de informações conhecidas pelo usuário, como, por exemplo, o uso de senha (uma sequência de caracteres previamente memorizada), sendo esta a forma de autenticação mais comumente utilizada;
- a partir dos atributos físicos ou das características biométricas do usuário.

Existem diversas técnicas biométricas que exploram diferentes partes e características físicas do corpo, como, por exemplo, a análise da face, da íris dos olhos, da digital dos dedos, dentre outras [15].

A dinâmica da digitação é uma técnica biométrica que busca identificar pessoas por meio do seu padrão de digitação [16]. Segundo essa técnica, na medida em que o usuário digita informações no teclado, são coletados dados em sua forma bruta (denominação adotada neste artigo para indicar dados que não tenham recebido nenhum tratamento especial). Tais dados são armazenados em uma estrutura simples, tal como a indicada na **tabela 1** e compreendem o código da tecla pressionada, o tempo (escrito como uma sequência de números que representa a quantidade de milissegundos desde primeiro de janeiro de 1970 [17] até o momento da ação), e a ação, que é o tipo de evento ocorrido.

Tab. 1 – Estrutura dos Dados Brutos

Tecla	Tempo	Ação
65	21469976126324	Down
65	21469976126430	Up
70	21469976126525	Down
70	21469976126629	Up

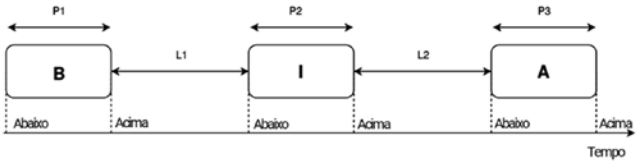

Fig 1 - Processo de coleta de dados da digitação. Fonte: [19].

Desses dados brutos são extraídas duas informações: latência, tempo entre as teclas pressionadas e tempo total de pressionamento de cada tecla.

O tempo de pressionamento corresponde ao tempo em que uma tecla demorou sendo pressionada. Para conseguir esta informação, basta diminuir o instante de tempo capturado no momento em que a tecla foi solta pelo instante de tempo capturado quando a tecla foi pressionada [18]. Por exemplo, na **figura 1** os eventos de pressionamento e soltura da tecla são representados por Abaixo (*Down*) e Acima (*Up*), respectivamente. Portanto, o resultado da subtração do tempo obtido no evento *Up* ocorrido na tecla "B" com o de *Down* da mesma é igual ao tempo de pressionamento, representado por P1. O mesmo se repete para as demais teclas.

A latência é o tempo entre o pressionamento de duas teclas. Ela pode ser explorada de diferentes maneiras. Por exemplo, podem ser utilizadas as distâncias entre dois eventos Down consecutivos de duas teclas, ou dois eventos *Up* consecutivos, entre outros. Alternativamente, pode ser utilizada a combinação de diferentes eventos [18]. No exemplo da **figura 1**, a latência é o tempo decorrido entre o evento *Up* de uma tecla com o evento *Down* da tecla seguinte, representada por L1 e L2. Diferentemente do tempo de pressionamento, a latência pode resultar em valores negativos, bastando uma tecla ser pressionada antes da tecla anterior a ela ser solta.

Em geral, após a coleta dos dados brutos, é executada uma fase de extração de características consideradas relevantes para o processo de reconhecimento do usuário [20]–[22]. Tal fase consiste em gerar novos atributos a partir dos dados coletados. A média e o desvio padrão do tempo de pressionamento das teclas são exemplos de informações que podem ser extraídas a partir da consolidação dos dados brutos coletados durante a digitação dos textos. Este tipo de consolidação permite construir modelos de reconhecimento que sejam independentes das teclas digitadas pelos usuários.

Uma providência comumente adotada após a extração de características dos dados coletados é a normalização dessas características [21]. A normalização de dados consiste em ajustar os valores dos atributos de forma que todos assumam uma mesma ordem de grandeza [23]. Tal providência é importante para evitar que alguns atributos, por apresentarem escalas de valores em ordem de grandeza maior que outros, influenciem de forma tendenciosa a construção dos modelos de reconhecimento de usuário.

Em geral, o reconhecimento de usuário é tratado como um problema de classificação, uma das mais importantes tarefas de aprendizado de máquina [23]. Existem diferentes algoritmos de classificação que podem ser utilizados no reconhecimento de usuário [24], inclusive no contexto da dinâmica da digitação [25]–[27]. Detalhes sobre esses algoritmos se encontram na próxima seção.

2.2 Aprendizado de máquina

No contexto de aprendizado de máquina, a tarefa de classificação consiste em, dado um conjunto de registros de entrada associados a classes (rótulos categóricos pré-definidos), construir um modelo que seja capaz de mapear novos registros nas classes pré-definidas correspondentes [23]. Em geral, a tarefa de classificação segue a abordagem de aprendizado supervisionado, que se divide em duas fases: uma de treinamento e outra de testes.

A fase de treinamento compreende a abstração de um modelo de conhecimento a partir dos dados apresentados na forma de pares ordenados (entrada,

saída desejada). No contexto de reconhecimento de usuários baseado na dinâmica da digitação, cada par ordenado corresponde a uma amostra dos dados coletados. A entrada consiste de um conjunto de valores extraídos e normalizados a partir dos dados brutos da amostra (vide Seção 2.1). A saída desejada é uma classe entre duas possíveis: positiva (caso o texto da amostra tenha sido digitado pelo usuário correspondente) ou negativa (caso contrário). Diferentes amostras devem ser utilizadas no conjunto de treinamento, sendo a quantidade dessas amostras, das duas classes, equilibrada.

A etapa de testes tem como objetivo avaliar o desempenho do modelo de classificação produzido pela etapa anterior. Para isso, recebe como entrada um conjunto de teste formado por pares ordenados (entrada, saída desejada) diferentes dos utilizados no treinamento. A entrada de cada par é submetida ao modelo que produz uma resposta e a compara com a saída desejada, computando acerto (caso a saída desejada coincida com a saída produzida) ou erro (caso contrário). Ao final desta etapa, o desempenho do modelo é aferido.

Dentre as medidas de desempenho usualmente utilizadas para avaliar modelos de classificação, a acurácia é uma das mais populares [23]. Ela informa o percentual de acertos do modelo em relação ao total de amostras do conjunto de testes. Logo, quanto maior a acurácia do modelo, melhor é o desempenho do referido modelo (deve ser considerada a proporção de dados entre as classes para isso).

A construção, a avaliação e a aplicação de modelos de classificação é realizada por meio de algoritmos de classificação [24].

Embora existam muitos algoritmos de classificação, nenhum deles é absolutamente melhor que os outros [24]. Assim, os algoritmos de classificação são geralmente avaliados e comparados entre si para verificar qual deles supera os demais em um dado contexto [23].

Uma breve descrição dos algoritmos utilizados neste trabalho é apresentada abaixo.

O *k-Nearest Neighbors* (k-NN) é um algoritmo de classificação baseado em distância. Para classificar um novo registro de dados, o k-NN calcula as distâncias entre um registro "n" e cada registro de dados do conjunto de treinamento. Em seguida, o algoritmo seleciona os registros de dados "n" mais próximos, identifica a classe mais frequente "c" entre o grupo selecionado e indica "c" como sua saída. A distância euclidiana é uma das métricas de distância mais comumente adotadas pelas implementações de k-NN.

No método do centroide mais próximo [28], cada classe é representada por um registro de dados do centroide (média de todos os registros de dados que pertencem à mesma classe no conjunto de treinamento). Dado um novo registro "r", o centroide mais próximo calcula iterativamente as distâncias entre "r" e cada registro centroide do conjunto de treinamento. Tem um funcionamento similar ao algoritmo K-NN, que calcula as distâncias entre todos os N elementos mais próximos para definir a classe. O algoritmo K-Means utiliza o método do centroide mais próximo, os pontos no espaço amostral são as médias de cada classe e assim é calculada a distância entre a média das classes. As distâncias euclidiana e de Manhattan são métricas usadas tradicionalmente pelas implementações do centroide mais próximo.

Árvores de decisão [29] são modelos de conhecimento representados por gráficos acíclicos. Em uma árvore de decisão, cada nó interno denota uma decisão sobre um atributo do conjunto de dados. Tal decisão determina como os registros de dados são particionados. Os nós das folhas indicam as classes. A escolha do atributo usado em cada decisão é geralmente guiada por índices como entropia ou *gini impurity* [30]. Existem alguns algoritmos que implementam Árvores de Decisão, por exemplo, ID3 [31], C4.5 [32], Árvore de Classificação e Regressão (ACR) [33], etc.

Random Forest é um algoritmo [34] que gera um conjunto de árvores de decisão baseado em amostras de dados selecionadas aleatoriamente do conjunto de treinamento. Dado um novo registro de dados "r", todas as árvores são aplicadas para classificar "r". O resultado de cada árvore de classificação é usado como voto para uma classe específica. A classe mais votada é designada como a classe resultante do algoritmo. A

velocidade é uma das principais vantagens da *Random Forest*.

A máquina de vetores de suporte (SVM) [35] usa classificadores lineares para criar dados ótimos que separam os hiperplanos. Para classificar registros de dados não lineares, o SVM mapeia os registros em um espaço dimensional mais alto, para que melhores partições de dados possam ser alcançadas. O mapeamento entre espaços é configurado por uma função do *kernel*. Linear, *radial basis function* (RBF) e Sigmoid são exemplos de funções alternativas do *kernel*.

O Classificador Bayesiano Ingênuo (NBC) é um algoritmo baseado em probabilidade condicional. Ele usa a Regra de Bayes, que assume independência de atributo para estimar a probabilidade de um determinado registro de dados de entrada ser atribuído a cada classe [36]. A classe com maior probabilidade é a saída da NBC.

Redes Neurais Artificiais são técnicas inspiradas no cérebro humano [37]. Uma rede neural é uma estrutura composta de componentes artificiais chamados neurônios. Basicamente, esses neurônios são geralmente organizados em camadas e as conexões entre eles são ponderadas. Durante a fase de treinamento da aprendizagem supervisionada, os pesos são ajustados para minimizar o erro entre a saída produzida pela rede e a saída desejada. Número de camadas, número de neurônios em cada camada são exemplos de parâmetros de redes neurais que devem ser configurados antes do treinamento. *Multilayer Perceptron* (MLP) é um popular modelo de rede neural com bons resultados em muitas aplicações [38].

A validação cruzada k-*fold* é uma técnica amplamente utilizada para comparar os desempenhos dos algoritmos de classificação [39]. Ele divide o conjunto de registros de dados em k subconjuntos chamados dobras. Uma "dobra" (um subconjunto dos dados, também conhecido como *folds*) é escolhida como o conjunto de teste e as dobras restantes (k-1) formam o conjunto de treinamento. Este processo é repetido ***k*** vezes. Cada iteração tem uma combinação diferente de conjuntos de teste e treinamento. O modelo aprendido é avaliado e seu desempenho é salvo. Após o loop, o desempenho médio do algoritmo em toda a iteração é calculado e salvo. O processo acima mencionado é repetido para cada algoritmo disponível. No final, os desempenhos médios dos algoritmos podem ser recuperados e comparados.

3. Trabalhos relacionados

3.1 Reconhecimento de usuários

É muito comum sistemas de informação serem projetados com apenas um processo de autenticação de usuários feito no momento do primeiro acesso e com base em um conjunto de caracteres (senha) que o usuário conhece [10]–[12]. Porém, nada impede que usuários informem suas credenciais a outras pessoas, o que compromete a fidedignidade do processo.

Diante deste cenário, existem diferentes técnicas de reconhecimento de usuários que exploram as características intrínsecas ao indivíduo, são as chamadas técnicas de biometria [26]

Entre as principais técnicas de biometria estão [15]:

- Reconhecimento da íris [40]: uma técnica que se baseia na extração de características da textura da íris. Apresenta boa acurácia, contudo necessita de equipamentos específicos para gerar imagens dos olhos e possui limitações quanto a movimentação da cabeça e da pálpebra;
- Reconhecimento facial [41], [42]: busca identificar pessoas por diferentes características ligadas a geometria da face e outras particularidades. Tem boa precisão, contudo necessita de dispositivos específicos para capturar imagens, além de ser intrusiva;
- Reconhecimento de voz [43]: identifica pessoas através do padrão de voz. Apresenta boa acurácia em ambientes controlados, contudo pode ter problemas com ruídos sonoros e distância do microfone, além de ser intrusiva;
- Reconhecimento de impressões digitais e de mãos [44], [45]: o reconhecimento de digitais é amplamente utilizado pois possui ótima acurácia apesar de necessitar de hardware específico.

Outra técnica biométrica de reconhecimento de usuário que vem ganhando notoriedade nos últimos anos é a dinâmica da digitação [27], [46]. É uma técnica que não demanda de hardware específico para coleta de dados e também não é intrusiva, ela é detalhada na próxima seção.

3.2 Reconhecimento de usuários baseado na dinâmica da digitação

A digitação de senhas periodicamente não é uma técnica viável de autenticação pois muitas aplicações não podem ter interrupções que distraiam os usuários. Adicionalmente, o uso de técnicas biométricas apresentadas nesta seção requer normalmente o uso de equipamentos caros e específicos.

O reconhecimento de usuário baseado na dinâmica da digitação é uma solução factível para mitigar esses problemas, pois é uma técnica biométrica. Tal característica impede que os usuários possam simplesmente passar suas credenciais para outras pessoas. Além disso, não necessita de nenhum equipamento adicional para a utilização.

De fato, a dinâmica da digitação para reconhecimento de usuários pode ser empregada de forma periódica e não intrusiva em sistema de informação [18], [47]. Cada vez que o usuário digita algo o sistema verifica a autenticidade do texto. A verificação pode ser feita sem o conhecimento do usuário, isso é importante em cenários onde o usuário não pode ser distraído. E por último, esta técnica não apresenta os potenciais riscos a privacidade que os métodos envolvendo reconhecimento facial, da íris e de voz.

As pesquisas que exploram a análise de padrões de digitação dividem-se em duas categorias: aquelas cujos padrões são identificados a partir da digitação de textos pré-definidos e aquelas que analisam padrões de digitação de textos dinâmicos (o usuário decide o que escrever) [16]. Em geral, o reconhecimento de usuários por meio de padrões de digitação pré-definidos apresenta melhor desempenho do que pela digitação de textos dinâmicos [48]. Isso ocorre basicamente porque os usuários escrevem sequências de caracteres previamente conhecidas, evitando que ocorram interrupções para pensamento ou consulta durante a digitação. O presente trabalho se enquadra na segunda categoria, na análise do padrão de digitação em textos dinâmicos.

Muitos estudos de análise de textos dinâmicos reportam uma alta acurácia (acima de 97%) em diferentes cenários com grande quantidade de usuários (100 ou mais) e amostras (pelo menos 2000) [16], [25], [49].

3.3 Mecanismos de autenticação em AVAs

Em AVAs, a autenticação de usuários baseada exclusivamente em usuário e senha propicia oportunidades de fraude, a partir do momento que pessoas não autenticadas podem facilmente substituir as autenticadas para executar tarefas, causando percepções incorretas sobre a performance acadêmica dos usuários [10].

A baixa confiabilidade dos sistemas baseados em usuário e senha tem encorajado o desenvolvimento de soluções alternativas, algumas delas vem do uso de técnicas biométricas [1], [12], [50]–[53].

Alguns estudos têm investigado o uso de ferramentas de reconhecimento facial, contudo, essa abordagem acaba violando a privacidade do usuário pois pode expô-lo em momentos íntimos [1], [12], [50], [52], [53]. Além disso, nenhum estudo real com grande volume de dados foi reportado por estes trabalhos.

Outros trabalhos combinaram diferentes tipos de biometria para o reconhecimento de usuários [12], [53]. No trabalho de [12], foi investigada uma abordagem baseada no reconhecimento de digitais e da face. No trabalho de [53], foi feito reconhecimento facial, e na análise dos padrões da digitação e do mouse. Embora a combinação de diferentes tipos de biometria possa melhorar a acurácia do processo de reconhecimento, ambos os trabalhos propuseram soluções que, além de violar privacidade, geram interrupções em diferentes momentos do uso do sistema, forçando o usuário a se autenticar diversas vezes ao longo do processo. Além disso, para viabilizar a abordagem proposta por esses trabalhos, é preciso que o usuário possua uma câmera, o que nem sempre ocorre em computadores pessoais ou empresariais. E também, nenhum estudo com grande volume de dados foi reportado nestes trabalhos.

Utilizar a dinâmica da digitação no reconhecimento de usuários EaD pode mitigar os problemas

mencionados. O processo de reconhecimento pode ocorrer de forma periódica e transparente para o usuário. E, além disso, o único requisito de hardware necessário é um teclado, componente usualmente necessário na utilização de AVAs em geral.

Apesar do exposto, não foram encontrados trabalhos que investigassem a viabilidade do uso da abordagem de reconhecimento de usuário baseada na dinâmica da digitação aplicada de forma periódica, não intrusiva e integrada a AVAs. Embora o mecanismo proposto em [1] produza bons resultados preliminares, a amostra avaliada é consideravelmente pequena (apenas 17 usuários). O presente trabalho é baseado nesta abordagem em um estudo de caso envolvendo 307 usuários, que produziram ao todo 4829 palavras e geraram mais de 1,6 milhão de registros de *keystrokes*.

4. Sistema de autenticação periódica proposto

Conceitualmente, o processamento do mecanismo proposto se divide em três fases. A primeira é responsável por coletar amostras de digitação de usuários do AVA e permitir a construção de um modelo de reconhecimento específico para cada usuário do ambiente. A segunda fase compreende a aplicação desses modelos na medida em que os respectivos usuários utilizam as funcionalidades do AVA. A cada vez que um usuário digita algum texto em uma funcionalidade do ambiente, o mecanismo proposto avalia o padrão de digitação, aplicando o modelo de reconhecimento correspondente, e armazena o resultado da avaliação junto ao texto gravado. A terceira fase compreende a apresentação de relatórios periódicos que indicam usuários responsáveis por ocorrências de digitação cuja autoria não tenha sido reconhecida pelo mecanismo proposto. As **figuras 1, 2 e 3** ilustram graficamente os módulos funcionais da primeira, segunda e terceira fases, respectivamente. As próximas subseções detalham cada um desses módulos funcionais.

Cabe ainda ressaltar que para que o mecanismo proposto opere de forma integrada ao AVA desejado, o administrador do sistema deve configurar o ambiente, definindo que funcionalidades deverão ser submetidas ao processo de autenticação periódica. É importante também notar que a análise da dinâmica da digitação é complementar ao reconhecimento inicial de usuário por meio de login e senha.

4.1 Coleta de dados

De forma a verificar se um usuário é realmente quem ele diz ser, usando a dinâmica da digitação, é necessário antes, capturar o padrão de digitação desse usuário. O módulo de Coleta de Dados, primeiro módulo da **figura 2**, tem como objetivo capturar amostras do padrão de digitação.

Com o intuito de identificar o padrão de digitação do usuário, o módulo de Coleta de Dados apresenta um formulário em branco com perguntas cujas respostas precisam ser digitadas. Em algumas delas, o conteúdo do texto a ser digitado é fixo. Nas demais o texto é dinâmico, podendo ser digitado livremente pelo usuário.

Na medida em que o usuário preenche o formulário, os registros de digitação de cada tecla são coletados de forma transparente ao usuário e são armazenados em uma espécie de histórico, que contém os dados brutos, no formato apresentado na **tabela 2**.

4.2 Construção do modelo de reconhecimento

Esta seção detalha o segundo módulo apresentado na **figura 2**. Nele, um modelo de reconhecimento é construído para cada usuário do AVA. A construção de cada modelo é um processo que ocorre em quatro etapas, sob coordenação do administrador do sistema.

Uma vez escolhido um usuário do AVA, a primeira etapa do processo de construção do modelo de reconhecimento é a extração de características da digitação (*keystrokes*, tempo de pressionamento das teclas e o tempo de latência entre o pressionamento sequencial de duas teclas) presente em cada amostra de texto.

Na segunda etapa, é feita a construção de atributos, em que tais características dão origem a novos atributos com informações consolidadas. Tais

atributos são: a média e o desvio padrão do tempo de pressionamento das teclas e da latência entre as teclas.

Na etapa seguinte, os atributos consolidados são normalizados, colocando-os na mesma ordem de grandeza para efeito de comparação e melhor aplicação no algoritmo responsável pelo treinamento.

Na última etapa, o Administrador do Sistema aplica diversos algoritmos de aprendizado de máquina a fim de selecionar qual o mais indicado para o usuário em questão. A comparação entre os algoritmos é feita por meio de um processo de validação cruzada em k-*folds*, sendo a acurácia, a métrica de avaliação escolhida. O critério quem guiou esta escolha foi o de que a acurácia pode compreender a precisão e a abrangência como uma métrica balanceada de distribuição de dados [24]. O algoritmo com o melhor desempenho médio na validação cruzada é considerado o vencedor e armazenado em uma base de conhecimento de forma a ser utilizado pelo módulo de aplicação do modelo de reconhecimento, sempre que for necessária a autenticação do usuário correspondente.

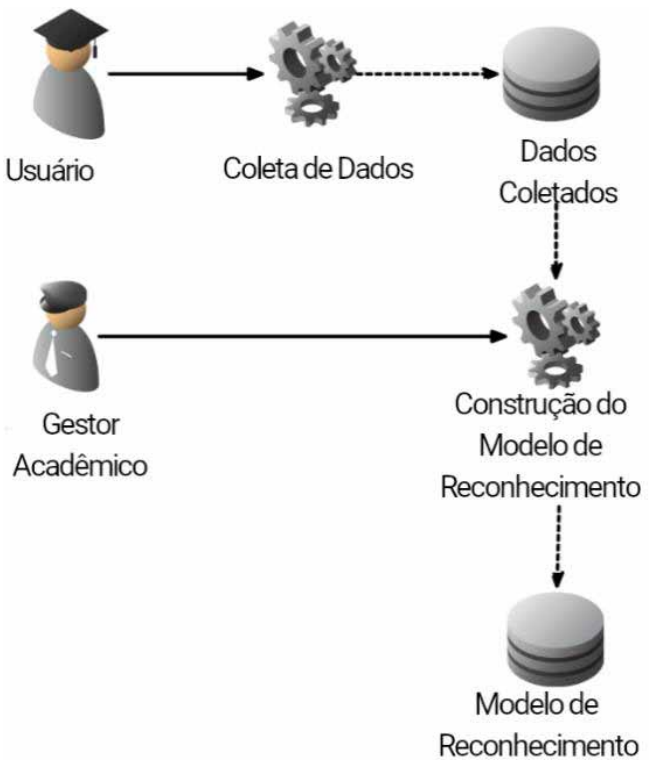

Fig. 2 – Primeira fase do mecanismo de autenticação periódica - Coleta de dados e construção do modelo de reconhecimento do usuário. Fonte: elaboração própria.

4.3 Verificação de autoria

Uma vez que o modelo de reconhecimento de um usuário tenha sido construído e armazenado na base de conhecimento, sempre que tal usuário acessar uma funcionalidade do AVA que esteja associada ao mecanismo proposto, o módulo de Verificação de Autoria é acionado (**figura 3**). Note que a funcionalidade do AVA não faz parte do mecanismo proposto, devendo ser ajustada para acionar o módulo de Verificação de Autoria.

O processamento deste módulo ocorre em três etapas. Durante a primeira etapa, os registros de digitação de cada tecla são coletados de forma transparente ao usuário, de forma análoga ao processo realizado pelo módulo de Coleta de Dados.

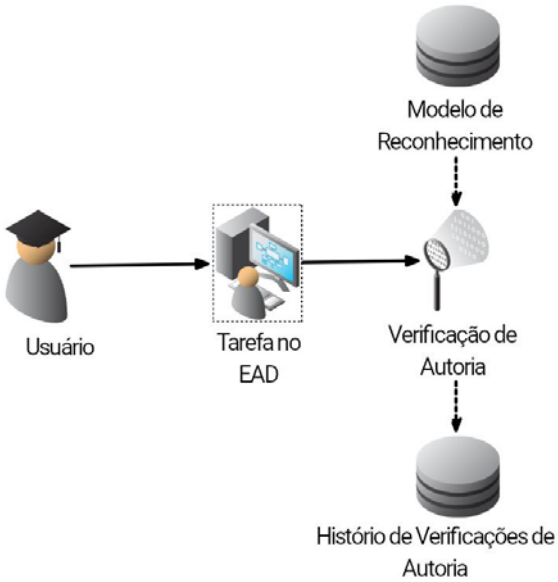

Fig 3 - Segunda Fase do Mecanismo Proposto: Verificação de Autoria. Fonte: elaboração própria.

Após o usuário concluir a digitação e os dados brutos serem salvos, a segunda etapa do módulo de Aplicação do Modelo de Reconhecimento é iniciada. De forma análoga ao descrito no módulo de Construção do Modelo Preditivo, nesta etapa, são calculados o tempo de pressionamento das teclas e o tempo de latência entre teclas. Em seguida é gerada a média e o desvio padrão de cada uma dessas características e por último os atributos são normalizados.

Na última etapa, o modelo de reconhecimento do usuário que está interagindo com o AVA é recuperado da base de conhecimento e aplicado sobre os dados processados na etapa anterior, a fim de verificar se o usuário em questão foi, de fato, o autor dos dados recém-digitados. Os resultados possíveis desta etapa são: autoria confirmada (classe positiva) ou autoria suspeita (classe negativa). O resultado obtido é armazenado de forma vinculada aos dados digitados e pode ser consultado pelo módulo de Análise de Autorias da terceira fase de operação do mecanismo proposto.

4.4 Análise de autorias

Na terceira e última fase de operação do mecanismo proposto, gestores acadêmicos podem acessar os resultados das verificações de autoria realizadas durante o período em que os usuários desenvolveram suas atividades no AVA. Isto é feito por meio do módulo de Análise de Autorias, representado na **figura 4**. Este módulo provê dois tipos de visão sobre os resultados: visão sintética e visão detalhada. Na visão sintética, o gestor acadêmico pode consultar os usuários do AVA pelos percentuais de autorias suspeitas e confirmadas dentro de um intervalo de tempo. Na visão detalhada, o gestor tem acesso às informações vinculadas a cada resultado, podendo visualizar, por exemplo, que tipo de funcionalidade foi acessada, qual o texto digitado pelo usuário, a data e a hora da digitação, dentre outras.

Fig. 4 - Terceira Fase do Mecanismo Proposto: Análise de Autorias. Fonte: elaboração própria.

[1] http://www.yahoo.com.br

O módulo de Análise de Autorias permite realizar diferentes análises longitudinais quanto à evolução dos percentuais de autorias suspeitas dos usuários. A decisão sobre como proceder junto a cada usuário deve ser norteada pela política pedagógica adotada em cada instituição/curso. Por exemplo, gestores acadêmicos podem optar por apenas acompanhar (sem nenhum contato), em novas atividades, usuários que tenham apresentado uma elevação pontual e isolada do percentual de autorias suspeitas em uma determinada atividade. Usuários com recorrentes e elevados percentuais de autoria suspeita podem ser contatados pelos gestores acadêmicos a fim que estes busquem um melhor diagnóstico das causas do problema e, em função dele, decidam como atuar em cada caso.

Outra questão importante que pode ser considerada sob o ponto de vista pedagógico nas análises é quanto à natureza das atividades e dos conteúdos associados a percentuais de autoria suspeita mais elevados. Atividades e conteúdos com elevada incidência de autorias suspeitas junto a diferentes usuários podem ser consequência do fato de que os recursos pedagógicos disponíveis sobre os respectivos assuntos no AVA estejam se mostrando inadequados e os usuários estejam recorrendo a apoio externo indevido. Diante deste cenário, uma revisão voltada ao aprimoramento de tais recursos pode ser uma intervenção recomendável.

Algoritmos devem ter a linha numerada e as palavras-chave devem estar em negrito.

5. Protótipo

A fim de demonstrar a viabilidade do mecanismo proposto neste artigo, optou-se por desenvolver um protótipo e integrá-lo a um AVA de código livre. Em função de sua popularidade, o AVA escolhido foi o Moodle.

Para extrair informações da digitação foi criada uma função de captura de dados em *javascript* (usando o YUI - *Yahoo! User Interface* (YUI, 2017))[1].

Esta função observa, com precisão de milissegundo, os eventos de quando o usuário aperta a tecla e de quando ele a solta. Esta função é utilizada nos módulos de coleta de dados e verificação de autoria. Cabe ressaltar ainda que tal função não sofre influência da latência da rede, pois opera em memória, na máquina do usuário. Isso diminui as chances de ocorrerem interferências no padrão de digitação do usuário.

Para processamento dos dados foram implementados serviços web em Python 3.5.2[2] com auxílio do micro *framework flask* 0.11[3]. Outrossim, foi utilizado a API *scikit-learn* versão 0.18 [54] para criar os modelos de reconhecimento. Para o armazenamento dos dados foi utilizado o sistema gerenciador de bancos de dados relacionais MySql 5.7.14[4].

5.1 Coleta de dados

A **figura 5** apresenta uma visão parcial da interface do módulo de coleta de dados. Cabe destacar que algumas respostas são solicitadas mais de uma vez, conforme recomendado em vários trabalhos relacionados à dinâmica da digitação [20], [21], [55].

A função de captura de dados foi vinculada a cada um dos campos a serem preenchidos pelo usuário. Os dados brutos coletados são em seguida, enviados para o servidor para serem armazenados no banco de dados.

Por força das características peculiares da topologia do mapa de rotas, o protocolo DD, um protocolo de roteamento multiponto, apresentou um desempenho, no que se refere à probabilidade de entrega apresentada na **tabela 2**, correspondente ao gráfico mostrado na **figura 2**.

Fig 5 - Formulário de Coleta de Dados. Fonte: elaboração própria.

5.2. Construção do modelo de reconhecimento

Para implementar o módulo de Construção do Modelo de Reconhecimento foram desenvolvidos dois subprogramas: o pré-processamento, que prepara os dados para serem utilizados pelos classificadores, e a construção do modelo do usuário, que compara diversos algoritmos de classificação com o intuito de selecionar o melhor modelo para o usuário.

Cumulativamente as funções descritas na Seção 4.2, o pré-processamento do subprograma implementou a extração de características (tempo de latências e pressionamento), a construção de atributos (média, variância e desvio padrão) e a normalização linear dos dados [56]. Os resultados foram gravados em um banco de dados, conforme ilustrado na **tabela 2**.

[2] http://www.python.org
[3] http://www.flask.pocoo.org
[4] http://www.mysql.com

Tab. 2 – Atributos Normalizados

id u	id pa	μ p	μ l	σ^2 p	σ^2 l	σ l	σ p
122	540	0.174	0.115	0.788	0.151	0.121	0.091
122	541	0.083	0.269	0.004	0.919	0.086	0.086
122	542	0.109	0.073	0.22	0.004	0.104	0.094
122	543	0.114	0.23	0.261	0.691	0.107	0.137
122	544	0.086	0.097	0.034	0.086	0.137	0.143

Legenda: **id** são os identificadores, **u** é o usuário e **pa** o texto utilizado, μ é a média, σ^2 a variância e σ o desvio padrão. **l** representa a latência e **p** o tempo de pressionamento.

No protótipo desenvolvido, o programa de construção do modelo do usuário não foi integrado ao Moodle, contudo o mesmo pode ser incluído em um produto final.

Os serviços foram codificados em linguagem Python e combinam funções disponíveis na biblioteca scikit-learn para executar a validação cruzada com os vários algoritmos apresentados na Seção 2.2 e identificar o melhor modelo de classificação gerado. Três saídas são produzidas, uma planilha com as acurácias dos modelos produzidos, um arquivo contendo o melhor modelo de classificação gerado para o usuário analisado e um registro de dados que é inserido em uma tabela do banco de dados denominada "modelo do usuário" que possui três campos, ID do usuário, algoritmo selecionado e diretório contendo o modelo do usuário.

5.3 Verificação de autoria

A funcionalidade do Moodle escolhida para ser integrada ao protótipo do mecanismo proposto foi o Fórum. Sua escolha se deve basicamente ao fato de ser um dos recursos mais usados do ambiente e que permite a obtenção de amostras de texto de tamanho e conteúdo variáveis. A função de captura de dados foi vinculada aos campos assunto e mensagem do formulário de criação de tópico de discussão mostradas na **figura 6** e também ao campo de postagem de respostas aos tópicos de discussão.

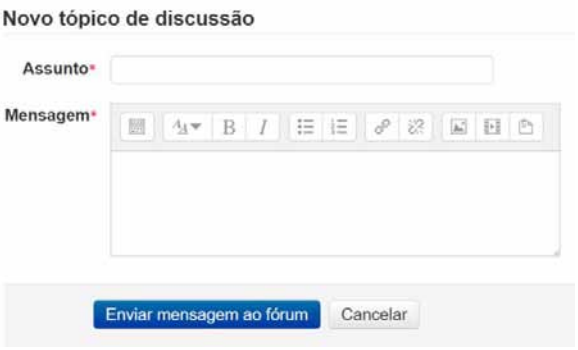

Fig 6 - Formulário de Postagem de Assuntos do Fórum – os conteúdos dos campos assunto e mensagem são coletados pelo protótipo e submetidos ao reconhecimento de usuário. Fonte: elaboração própria.

Assim sendo, sempre que um usuário digita algum conteúdo em um desses campos, o padrão de digitação é coletado e armazenado. Depois disso, um programa em Python é usado para habilitar a autenticação do usuário. Primeiro, é recuperado o modelo de classificação estabelecido para o usuário em questão, depois este é aplicado no texto selecionado e o mesmo é classificado em confirmado ou como tendo autoria suspeita. Finalmente, os resultados são salvos em uma tabela intitulada "histórico de verificação de autoria".

Cabe ressaltar que não foi necessário realizar nenhuma mudança nas interfaces originais do Moodle, sendo transparente para o usuário final. Desta forma, o uso do mecanismo proposto se comporta de forma não intrusiva durante seu processamento.

5.4 Análise de autoria

O propósito deste módulo consiste basicamente da geração de relatórios extraídos a partir de consultas realizadas sobre os resultados das verificações de autoria previamente executadas. As **figuras 7 e 8** ilustram dois desses relatórios. O relatório da **figura 7** apresenta os usuários em ordem decrescente de autorias suspeitas detectadas pelo sistema. Caso o gestor acadêmico deseje ver os detalhes dessas autorias, o relatório da **figura 8** pode ser consultado.

Gestor Acadêmico: Gestor Teste
Data do relatório: 31/07/2016
Período consultado: 1/07/2016 até 31/07/2016

Usuário	% de Autorias Certificadas	% de Autorias Suspeitas
B	50%	50%
C	90%	10%
A	99%	1%

Fig 7 - Exemplo de Relatório – Distribuição dos Resultados de Verificação de Autoria. Fonte: o autor.

Usuário: B
% de autenticações suspeitas: 50%

Funcionalidade	Referência	Data e Hora	Texto Digitado
Fórum - Novo tópico de discussão	Título	13/07/2016 : 12:32	Inteligência Artificial
Fórum - Novo tópico de discussão	Conteúdo	13/07/2016 : 12:32	Inteligência Artificial é uma área do conhecimento que estuda ...
Fórum - Resposta	Conteúdo	14/07/2016 : 13:00	As Redes Neurais podem ter várias camadas intermediáras e ...

Fig 8 - Exemplo de Relatório – Detalhamento das Autenticações Suspeitas Associadas a um Usuário. Fonte: elaboração própria.

6. Estudo de caso

Este estudo avaliou a viabilidade e qualidade do sistema de autenticação periódica proposto. Ele foi realizado em uma instituição que utiliza o Moodle como suporte a atividades pedagógicas. Foram realizadas coletas de 307 usuários, sendo 2 professores e 305 estudantes. O estudo realizado foi do estilo duplo-cego, ou seja, nem os participantes e aplicadores sabiam do processo de coleta envolvido.

Foram coletados um total de 3219 palavras e armazenados mais de 1,6 milhões de registros de teclas pressionadas e liberadas. A coleta foi realizada no mês de Julho de 2017.

Depois de coletados os dados, sete algoritmos de classificação (descritos na sessão 2.2) foram aplicados aos dados. Foram eles: K-NN, Nearest Centroid, SVM, Naive Bayes, Decision Tree, Random Forest and Artificial Neural Networks. A **tabela 3** apresenta as configurações utilizadas nos algorítimos.

[5] Algoritmo de classificação e regressão de em árvores de decisão

Tab. 3 – Algoritmos de Classificação

Algorithm	Configuration
K-NN	K=3; Distance = Euclidean
Nearest Centroid	Distance = Manhattan
SVM	Kernel = RBF; error penalty rate = 100;
Naive Bayes	NA
Decision Tree	split quality index = gini; split strategy = best split
Random Forest	number of trees = 13
Multi-Layer Perceptron	1 hidden layer with 200 units; lbfgs; epochs = 10,000

Todas as árvores de decisão foram produzidas com o algoritmo ACR[5].

Para cada usuário, cada algoritmo foi avaliado em um processo de validação cruzada com 3 divisões. O modelo de reconhecimento com a maior acurácia para cada usuário foi escolhido para realizar a validação dos seus textos futuros. Assim, este estudo de caso produziu 6447 modelos de reconhecimento (307 usuários x 7 algoritmos x 3 *folds*) e foi selecionado um por usuário.

A **tabela 4** mostra a média de performance no processo de validação cruzada de cada algoritmo em dez usuários selecionados aleatoriamente. As melhores acurácias estão em negrito e indicam o modelo que foi escolhido para o usuário na linha em questão, a acurácia gera valores de 0 a 100, esses valores podem ser considerados como percentuais de acerto da classe verdadeira, 100 seria um acerto total de todas as predições. A **figura 9** sumariza a distribuição de frequências de performances de cada modelos de reconhecimento selecionado.

Tab. 4 – Acurácia média de cada algoritmo em uma amostra de 10 usuários

Usuários	K-NN	Centroid	Dec. Tree	SVM	Bayes	R. Forest	MLP
6	75,00	50,00	**80,00**	55,00	75,00	70,00	60,00
51	83,33	85,19	81,48	53,70	**90,74**	87,04	81,48
60	59,62	32,69	38,46	46,15	**63,46**	55,77	50,00
65	78,00	52,00	**88,00**	52,00	74,00	80,00	62,00
77	95,45	95,45	**100,00**	95,45	95,45	95,45	95,45
83	**94,23**	67,31	80,77	65,38	80,77	78,85	86,54
123	**78,13**	71,88	75,00	68,75	56,25	75,00	59,38
134	56,25	54,17	85,42	66,67	75,00	**87,50**	70,83
242	92,86	82,14	96,43	60,71	67,86	**100,00**	92,86
247	91,67	91,67	**100,00**	91,67	95,83	91,67	75,00

Legenda: MLP – Rede Neural Multi-layer Perceptron

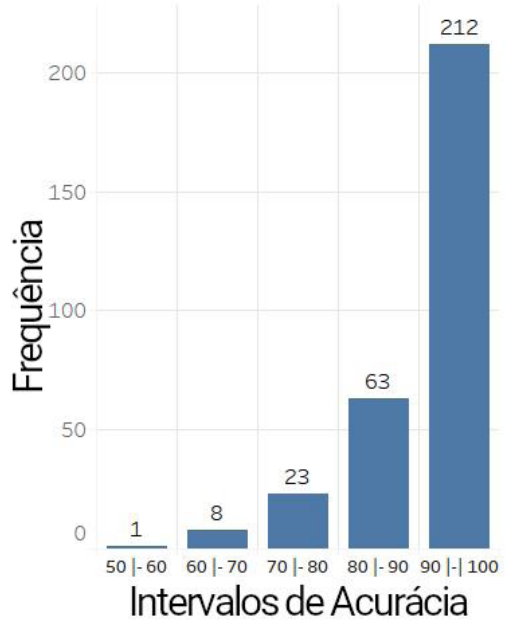

Fig 9 - Distribuição de frequências para a acurácia dos modelos de reconhecimento de usuários. Fonte: elaboração própria.

A **figura 9** mostra que 212 modelos de reconhecimento apresentaram acurácia entre 90% e 100%. Aproximadamente, 89% de todos os modelos apresentaram acurácia acima de 80%, indicando que o mecanismo permaneceu eficaz na identificação de autoria suspeita de textos e que pode ser uma ferramenta valiosa para apoiar decisões de gerenciamento acadêmico sobre má conduta.

Depois da construção do modelo, a fase de verificação de Autoria (seção 4.3) iniciou. A coleta foi feita em agosto de 2017 e foram inseridas mais de 1600 palavras. O mecanismo de autenticação proposto verificou a autenticidade de cada texto inserido pelo usuário. Dois professores acessaram o módulo para verificar os resultados de verificação de autoria.

A **figura 10** mostra, de forma ilustrativa, uma visão sintética de alguns resultados. É observado que altas ou baixas porcentagens de autorias suspeitas refletem o padrão geral dos usuários. Como não haveria razão para supor que outros usuários estivessem utilizando o sistema no lugar dos autenticados, pois a coleta foi controlada, propositalmente quatro usuários deixaram que outras pessoas usassem suas contas (ids **36, 173, 187 e 28**). Sendo esperado dessa forma que os resultados para esses usuários fossem de alta porcentagem de autoria suspeita.

Autorias

Usuário	%Certificadas	%Suspeitas
36	1.62	98.38
173	10.70	89.33
187	17.24	82.76
28	18.60	81.40
171	66.67	33.33
72	67.86	32.14
60	69.23	30.77
129	69.23	30.77
264	69.23	30.77
291	69.23	30.77
106	70.83	29.17
103	71.43	28.57
219	71.43	28.57
154	73.33	26.67
165	73.33	26.67
235	73.33	26.67

1 2 3 4 5 ... »

Fig 10 - Exemplo de relatório - Distribuição do Resultado da Verificação de Autoria. Fonte: elaboração própria.

Com o objetivo de avaliar as usabilidades das interfaces e os potenciais usos pedagógicos da ferramenta, duas análises foram feitas, baseadas em uma pesquisa qualitativa realizada com os usuários após usarem o AVA. As questões nesta pesquisa foram baseadas na heurística de usabilidade da interface proposta, considerando as experiências anteriores em Moodle pelo publico alvo. O questionário foi composto de questões fechadas com três alternativas cada, de acordo com a escala de *likert*[6].

Quase todos os usuários responderam questões sobre a ferramenta de verificação de autoria, apenas dois usuários responderam sobre a ferramenta de análise de autoria. Deve-se notar que, embora o módulo de verificação de autoria não seja acessado diretamente pelos usuários, ele foi incluído na enquete porque foi integrado ao Fórum (funcionalidade

[6] é um tipo de escala de resposta psicométrica usada habitualmente em questionários.

Moodle usada no estudo de caso). Deve-se notar também que, no caso do módulo de análise de autoria, questões abertas foram adicionadas à enquete de modo a permitir que os professores expressassem suas opiniões sobre o protótipo e as possibilidades de ações pedagógicas oferecidas pelo módulo.

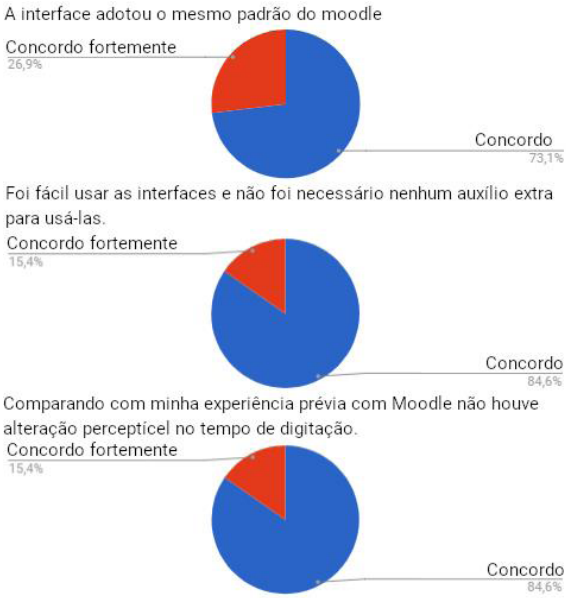

Fig 11 - Análise da Interface de Coleta de Dados. Fonte: elaboração própria.

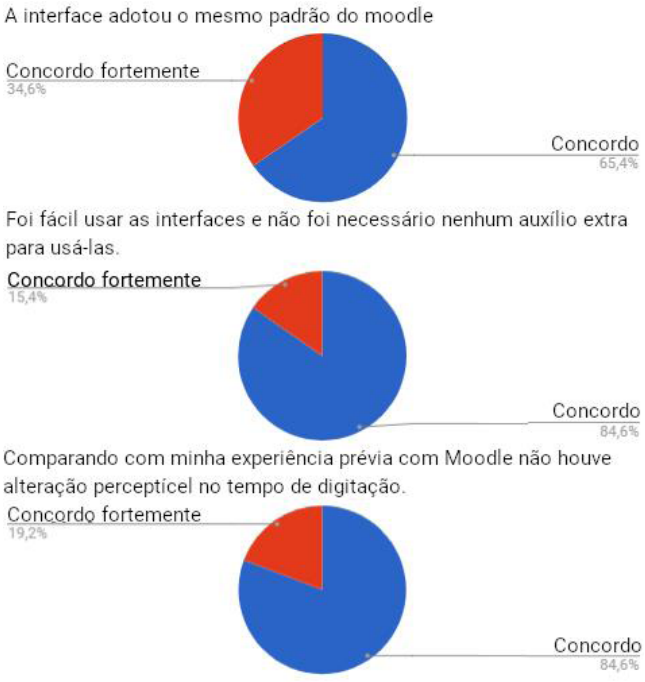

Fig 12 - Análise da interface de verificação de autoria. Fonte: elaboração própria.

Todos os usuários concordaram que as interfaces dos módulos de coleta de dados e verificação de autoria seguiram o mesmo padrão do Moodle. De fato, a compatibilidade das interfaces do protótipo com o AVA foi uma das premissas adotadas na implementação. Da mesma forma, os entrevistados afirmaram que era fácil entender como usar os módulos. E, por fim, os usuários não notaram diferença no tempo de resposta entre o uso das interfaces com a verificação de autoria e as interfaces originais do Moodle.

Em relação às respostas dos professores sobre as questões abertas, ambos concordaram que o módulo de Análise de Autoria apresenta capacidade para identificar possíveis problemas de autenticidade. Em sua opinião, o protótipo do módulo não é suficiente para sustentar a viabilidade de uma política pedagógica de combate à fraude relacionada à autoria de atividades a distância. Segundo eles, para poder suportar tal política, novas características devem ser desenvolvidas. Um exemplo interessante citado por um corpo docente foi a implementação de um canal de comunicação entre gestores acadêmicos e usuários de AVA para lidar com questões relacionadas à autoria de atividades on-line desenvolvidas na plataforma. Por meio do uso desse canal, uma memória das interações entre gerentes e usuários, além de registros sobre as ações realizadas em cada caso, seria construída. No futuro, essa memória pode ser usada para aprimorar os modelos de reconhecimento de usuários do AVA.

Outra sugestão importante foi a introdução dos conceitos de certificação automática de autoria e certificação manual de autoria. O primeiro seria utilizado para as certificações realizadas automaticamente pelo sistema, enquanto o segundo seria informado pelos gestores acadêmicos como consequência de seus contatos com os usuários. Esse tipo de informação também pode oferecer suporte futuro para melhorar os modelos de reconhecimento do sistema.

7. Conclusão

Um dos maiores desafios da educação à distância é a autenticidade do usuário durante o período de

atividade em um AVA. A autenticação em AVAs são geralmente restritas ao momento em que o usuário informa uma senha para se conectar ao ambiente. O problema com essa abordagem é que os usuários não credenciados, após a autenticação inicial, podem assumir o papel de usuários credenciados, o que pode levar a vários problemas de segurança e pedagógicos. Embora a autenticação periódica seja uma alternativa para atenuar esses problemas, técnicas biométricas, como reconhecimentos facial, de voz e de íris, que acabam sendo soluções muito intrusivas.

Neste contexto, um mecanismo que realiza autenticação periódica não intrusiva em AVAs foi proposto pela primeira vez por [1]. Ele usa técnicas de aprendizado de máquina para construir modelos de reconhecimento com base na dinâmica de pressionamento de tecla dos usuários e pode ser integrado em diferentes plataformas de AVA existentes. Embora este motor tenha apresentado bons resultados preliminares, foi avaliado em um cenário com apenas 17 usuários. Assim, a fim de demonstrar a viabilidade prática do motor em um cenário real com mais usuários e dados, este artigo relatou um estudo de caso no qual o mecanismo foi integrado ao Moodle e aplicado a um grupo de 307 usuários que produziram 4.829 palavras avaliadas pelos modelos de reconhecimento. Mais de 1,6 milhão de toques foram coletados e processados. A ferramenta gerou modelos com precisão superior a 80% para 89% dos usuários. Uma pesquisa qualitativa confirmou que a usabilidade da interface do protótipo era adequada. O estudo de caso apresentou evidências práticas de que o mecanismo de autenticação pode ser uma ferramenta valiosa para apoiar decisões de gerenciamento acadêmico sobre má conduta mesmo em cenários com muitos usuários e dados.

São exemplos de possíveis ações a serem tomadas: ofertas de reforço pedagógico direcionado aos conteúdos das atividades associadas às autorias suspeitas, possivelmente incluindo novas atividades a serem desenvolvidas, revisão técnica voltada ao ajuste ou mesmo à substituição do modelo de reconhecimento de usuário, notificação formal ao usuário quanto à necessidade de lisura do processo acadêmico.

Como trabalho futuro, pretendemos investigar como melhorar o desempenho dos modelos de reconhecimento. Além disso, planejamos adicionar novos recursos ao módulo de análise de autoria para que ele possa suportar as decisões dos gerentes acadêmicos e refinar o protótipo para fornecer um plug-in de código aberto para o Moodle.

Referências Bibliográficas

[1] M. A. S. Cruz, J. C. Duarte, e R. R. Goldschmidt, “Keystroke Dynamics Applied to Periodic Authentication in Virtual Learning Environments”, Brazilian J. Comput. Educ., 2017.

[2] R. C. Clark e R. E. Mayer, E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. John Wiley & Sons, 2016.

[3] “Censo da Educação Superior, Notas Estatísticas”. 2014.

[4] M. Mandaji, “O processo de colaboração nos trabalhos de coautoria em ambientes virtuais de aprendizagem”, Rev. Bras. Informática na Educ., vol. 20, no 1, p. 50, 2012.

[5] F. P. Galafassi, J. C. Gluz, e C. Galafassi, “Análise crítica das pesquisas recentes sobre as tecnologias de objetos de aprendizagem e ambientes virtuais de aprendizagem”, Rev. Bras. Informática na Educ., vol. 21, no 3, p. 41–52, 2013.

[6] M. A. M. Bucci e P. da Silva Meneghel, “Tecnologias e ferramentas gratuitas da Internet e sua aplicação aos programas de aprimoramento profissional à distância de equipes em bibliotecas universitárias”, RBBD. Rev. Bras. Bibliotecon. e Doc., vol. 4, no 2, p. 52–63, maio 2009.

[7] A. M. Poersch, N. S. Santos, e M. A. V Nelson, “Estudo quantitativo da manutenção evolutiva em dois sistemas de código aberto”, in II Workshop de Manutenção de Software Moderno, 2006.

[8] D. L. King e C. J. Case, “E-Cheating: Incidence and trends among college students.”, Issues Inf. Syst., vol. 15, no 1, 2014.

[9] J. Moten Jr, A. Fitterer, E. Brazier, J. Leonard, e A. Brown, “Examining Online College Cyber Cheating Metho-

ds and Prevention Measures.", Electron. J. E-learning, vol. 11, no 2, p. 139–146, 2013.
[10] K. Rabuzin, M. Baca, e M. Sajko, "E-learning: Biometrics as a Security Factor", in 2006 International Multi-Conference on Computing in the Global Information Technology - (ICCGI'06), 2006, p. 64.
[11] E. Marais, D. Argles, e B. von Solms, "Security Issues Specific to e-Assessments", 8th Annu. Conf. WWW Appl., 2006.
[12] A. Moini e A. M. Madni, "Leveraging Biometrics for User Authentication in Online Learning: A Systems Perspective", IEEE Syst. J., vol. 3, no 4, p. 469–476, 2009.
[13] G. E. Violettas, T. L. Theodorou, e G. C. Stephanides, "E-Learning Software Security: Tested for Security Vulnerabilities & Issues", in 2013 Fourth International Conference on e-Learning "Best Practices in Management, Design and Development of e-Courses: Standards of Excellence and Creativity", 2013, p. 233–240.
[14] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication", Proc. IEEE, vol. 91, no 12, p. 2021–2040, dez. 2003.
[15] A. K. Jain, K. Nandakumar, e A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities", Pattern Recognit. Lett., vol. 79, p. 80–105, ago. 2016.
[16] A. Alsultan e K. Warwick, "Keystroke dynamics authentication: a survey of free-text methods", Int. J. Comput. Sci. Issues, vol. 10, no 4, p. 1–10, 2013.
[17] E. Ecmascript, "Language Specification". 2015.
[18] S. P. Banerjee e D. Woodard, "Biometric Authentication and Identification Using Keystroke Dynamics: A Survey", J. Pattern Recognit. Res., vol. 7, no 1, p. 116–139, 2012.
[19] G. C. Boechat, J. C. Ferreira, e E. C. B. Carvalho Filho, "Authentication personal", in 2007 International Conference on Intelligent and Advanced Systems (ICIAS), 2007, p. 254–256.
[20] C. Costa, G. F. Yared, R. N. Rodrigues, J. B. Yabu-Uti, F. Violaro, e L. L. Ling, "Autenticação Biométrica via Dinâmica da Digitação em Teclados Numéricos", in XXII Simpósio Brasileiro de Telecomunicações--SBrT'05, 2005, p. 4–8.
[21] G. Cavalcanti, "Composição de biometria para sistemas multimodais de verificação de identidade pessoal", Universidade Federal de Pernambuco, 2005.
[22] L. C. F. Araújo, L. H. R. Sucupira, M. G. Lizarraga, L. L. Ling, e J. B. T. Yabu-Uti, "User authentication through typing biometrics features", IEEE Trans. signal Process., vol. 53, no 2, p. 851–855, 2005.
[23] R. Goldschmidt, E. Bezerra, e E. Passos, Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações. Elsevier, Rio de Janeiro, 2015.
[24] K. Faceli, A. C. Lorena, J. Gama, e A. Carvalho, Inteligência Artificial: Uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.
[25] S. Bhatt e T. Santhanam, "Keystroke dynamics for biometric authentication—A survey", in Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on, 2013, p. 17–23.
[26] F. Monrose e A. D. Rubin, "Keystroke dynamics as a biometric for authentication", Futur. Gener. Comput. Syst., vol. 16, no 4, p. 351–359, fev. 2000.
[27] P. S. Teh, A. B. J. Teoh, e S. Yue, "A Survey of Keystroke Dynamics Biometrics", Sci. World J., vol. 2013, no 4, p. 1–24, nov. 2013.
[28] R. Tibshirani, T. Hastie, B. Narasimhan, e G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression", Proc. Natl. Acad. Sci., vol. 99, no 10, p. 6567–6572, maio 2002.
[29] J. R. Quinlan, C4. 5: programs for machine learning. Elsevier, 2014.
[30] U. M. Fayyad e K. B. Irani, "The attribute selection problem in decision tree generation", in AAAI, 1992, p. 104–110.
[31] J. R. Quinlan, "Induction of decision trees", Mach. Learn., vol. 1, no 1, p. 81–106, 1986.
[32] J. R. Quinlan, C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
[33] L. Breiman, J. H. Friedman, R. A. Olshen, e C. J. Stone, "Classification and regression trees (CART) Wadsworth International Group", Belmont, CA, USA, 1984.
[34] L. Breiman, "Random forests", Mach. Learn., vol. 45, no 1, p. 5–32, 2001.
[35] C. Cortes e V. Vapnik, "Support-vector networks", Mach. Learn., vol. 20, no 3, p. 273–297, 1995.
[36] P. Langley, W. Iba, e K. Thompson, "An analysis of Bayesian classifiers", AAAI, 1992.
[37] F. Rosenblatt, "Principles of neurodynamics", 1962.
[38] T. Kohonen, "An introduction to neural computing", Neural Networks, vol. 1, no 1, p. 3–16, 1988.

[39] R. Kohavi e others, "A study of cross-validation and bootstrap for accuracy estimation and model selection", in Ijcai, 1995, vol. 14, no 2, p. 1137–1145.
[40] J. Daugman, "How Iris Recognition Works", IEEE Trans. Circuits Syst. Video Technol., vol. 14, no 1, p. 21–30, jan. 2004.
[41] O. M. Parkhi, A. Vedaldi, e A. Zisserman, "Deep Face Recognition.", BMVC, 2015.
[42] R. Jafri e H. R. Arabnia, "A survey of face recognition techniques.", Jips, vol. 5, no 2, p. 41–68, 2009.
[43] J. Padmanabhan e M. J. J. Premkumar, "Machine Learning in Automatic Speech Recognition: A Survey", IETE Tech. Rev., vol. 32, no 4, p. 240–251, fev. 2015.
[44] M. M. H. Ali, V. H. Mahale, P. Yannawar, e A. T. Gaikwad, "Overview of fingerprint recognition system", in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, p. 1334–1338.
[45] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, e A. K. Jain, "FVC2000: Fingerprint verification competition", IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no 3, p. 402–412, 2002.
[46] Y. Zhong e Y. Deng, "A survey on keystroke dynamics biometrics: approaches, advances, and evaluations", Recent Adv. User Authentication Using Keystroke Dyn. Biometrics. Sci. Gate Publ., p. 1–22, 2015.
[47] P. S. Dowland, S. M. Furnell, e M. Papadaki, "Keystroke Analysis as a Method of Advanced User Authentication and Response", in Security in the Information Society, Boston, MA: Springer US, 2002, p. 215–226.
[48] D. Gunetti e C. Picardi, "Keystroke analysis of free text", ACM Trans. Inf. Syst. Secur., vol. 8, no 3, p. 312–347, ago. 2005.
[49] M. Curtin et al., "Keystroke biometric recognition on long-text input: A feasibility study", Proc. Int. MultiConf. Eng. Comput. Sci., 2006.
[50] F. A. Diniz, F. M. M. Neto, F. das Chagas Lima Júnior, e L. M. de O Fontes, "RedFace: Um Sistema de Reconhecimento Facial para Identificação de Estudantes em um Ambiente Virtual de Aprendizagem", RENOTE, vol. 10, no 3, dez. 2012.
[51] B. E. Penteado e A. N. Marana, "Autenticação biométrica on-line de usuários em aplicações web de Ensino a distância", in Companion the XIV Brazilian Symposium, 2008, p. 53.
[52] A. L. Rolim e E. P. Bezerra, "Um sistema de identificação automática de faces para um ambiente virtual de ensino e aprendizagem", in Companion the XIV Brazilian Symposium, 2008, p. 129.
[53] M. K. Dehnavi, S. M. Sharafi, e N. Nematbakhsh, "DEVELOPING A E-LEARNING MODEL FOR TRACKING THE CONTINUOUS ATTENDANCE OF THE STUDENTS.", J. Theor. Appl. Inf. Technol., vol. 24, no 1, 2011.
[54] F. Pedregosa et al., "Scikit-learn: Machine Learning in {P}ython", J. Mach. Learn. Res., vol. 12, p. 2825–2830, 2011.
[55] A. Darabseh e A. S. Namin, "The accuracy of user authentication through keystroke features using the most frequent words", in the 9th Annual Cyber and Information Security Research Conference, 2014, p. 85–88.
[56] J. Han, J. Pei, e M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.