

## RESUMO

A cada dia cresce a utilização de computadores e celulares para acessar informações confidenciais como contas bancárias, realizar atividades críticas como compras online com cartão de crédito. Em todos estes casos, a autenticação do usuário é necessária. Entretanto, sabe-se que a maior parte dos usuários utiliza senhas não-seguras e/ou as reutiliza em diversos outros serviços.

O objetivo deste trabalho é amenizar os problemas de autenticação pelo uso de senhas a partir de técnicas de inteligência artificial para criar um modelo de aprendizado de máquina capaz de identificar e autenticar um usuário a partir de características comportamentais referentes a digitação, permitindo (ou não) o acesso ao sistema.

## COLETA DE DADOS

A primeira parte do trabalho consistiu em uma coleta de dados. Para isso, foi desenvolvido um aplicativo para celular no qual o usuário deveria responder 3 perguntas dissertativas. O usuário deveria escrever por volta de 5 minutos em cada sessão. Foram realizadas 10 sessões com 3 voluntários e 1 sessão com outros 4 voluntários.

Durante as sessões, foram coletadas as leituras do acelerômetro, giroscópio e magnetômetro do dispositivo; além dos dados referentes aos toques.

## PROCESSAMENTO DOS DADOS E APRENDIZADO

Com os dados coletados, seguiu-se para a segunda parte do trabalho que foi composta por 3 etapas:

1. União dos dados coletados com um conjunto de dados externo;
2. Processamento dos dados em janelas para construção de *features*;
3. Uso destes dados para o treinamento dos modelos de aprendizado de máquina.

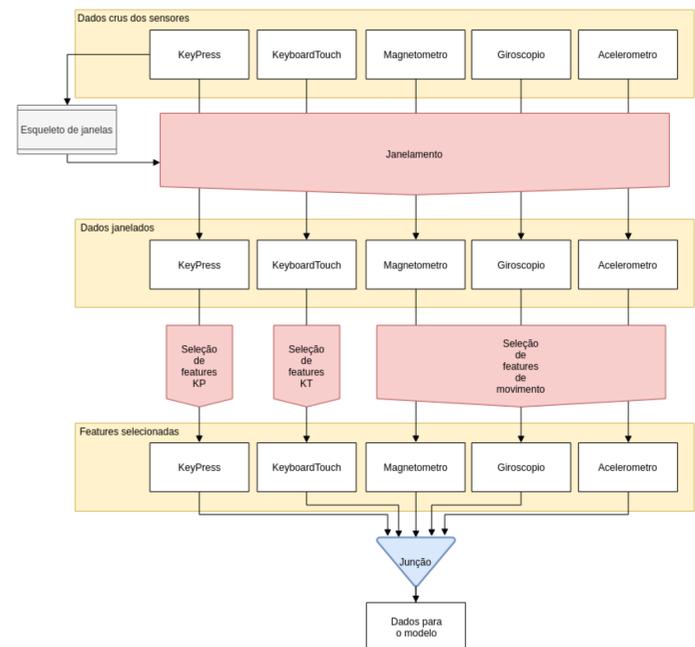


Diagrama explicando o processamento de dados

Foi treinando um classificador *Xgboost* para cada um dos voluntários com 10 sessões

## TESTE

Para se testar os classificadores, desenvolveu-se uma nova área do aplicativo no qual o usuário opta por entrar na conta de um dos usuários e deve responder uma pergunta para isto. Desenvolveu-se um servidor em *Django* que recebia os dados coletados de teste e respondia à requisição com a previsão do modelo daquela sessão ter sido feito pelo usuário legítimo. Para esta etapa, mais de 20 voluntários realizaram 3 sessões cada.

## RESULTADOS

Os modelos foram avaliados a partir da métrica *F-Score*, com um limiar de 95% para uma sessão ser classificada como legítima, dada a gravidade de um falso positivo no caso de uso de autenticação de usuários.

Modelo	Precisão	Revocação	<i>F0.5-Score</i>	<i>F2-Score</i>	<i>F1-Score</i>
1	1.0	1.0	1.0	1.0	1.0
2	0.67	0.8	0.69	0.77	0.73
3	1.0	0.6	0.88	0.65	0.75

**Integrantes:** Diego Amaral Souza  
Jonathan Yuiti Suzuki  
Thiago Cordeiro da Fonseca

**Professor Orientador:** Dr<sup>a</sup> Anarosa Alves Franco Brandao