

UNIVERSIDADE FEEVALE

LUIS MIGUEL PIRES

APLICAÇÃO DE MACHINE LEARNING PARA DETECÇÃO DE
PADRÕES NO COMPORTAMENTO DE USUÁRIOS EM UM
SISTEMA WEB

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2019

LUIS MIGUEL PIRES

APLICAÇÃO DE MACHINE LEARNING PARA
DETECÇÃO DE PADRÕES NO COMPORTAMENTO DE
USUÁRIOS EM UM SISTEMA WEB

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientador: Gabriel da Silva Simões

Novo Hamburgo
2019

RESUMO

Recentemente o modelo de entrega de *softwares* denominado *Software as a Service* (SaaS) têm ganhado muita popularidade, contando com uma previsão de crescimento de 21.2% ao ano entre 2018 e 2023. Essa rápida expansão gera um grande volume de horas necessárias com treinamentos à novos usuários e suporte técnico aos clientes, gerando um alto custo com equipes e dificultando a competitividade das empresas desse setor no mercado. Existem serviços que visam facilitar a adaptação de novos usuários em sistemas *online* e reduzir custos com suporte das empresas fornecedoras de *softwares*, contudo, estes sistemas necessitam mão de obra técnica para serem configurados e podem apresentar inconsistências ao longo do tempo de vida do *software*. Com base nesse cenário, o presente trabalho buscará descobrir padrões de comportamento dos usuários já ativos em um sistema através de aprendizado de máquina, para assim prever possíveis ações a serem tomadas em uma determinada tela por novos usuários. Sendo assim, será apresentado qual a acurácia que os modelos podem ter ao determinar ações que os usuários devem tomar em uma tela e então, possibilitar a automatização do processo de criação de guias interativos.

Palavras-chave: *machine learning, software as a service, user onboarding.*

SUMÁRIO

MOTIVAÇÃO	4
OBJETIVOS	9
METODOLOGIA	10
CRONOGRAMA	11
BIBLIOGRAFIA	12

MOTIVAÇÃO

A utilização de sistemas de informação para gerenciamento de empresas vêm crescendo a cada dia, acompanhando a facilidade de acesso aos computadores e à internet no Brasil. De acordo com pesquisa realizada pela Fundação Getúlio Vargas (FGV), a base instalada de computadores no Brasil teve uma evolução de 9% em 2018 com um total de 177,5 milhões de computadores em uso neste período (MEIRELLES, 2019).

Em paralelo a isso, diversos modelos de licenciamento de *softwares* vêm sendo adotados ao longo dos anos e, recentemente, o *Software as a Service* (SaaS) tem ganhado popularidade, com uma previsão de crescimento à uma taxa anual de 21.2% entre os anos de 2018 e 2023 (BUSINESS WIRE, 2018). No modelo SaaS, o cliente paga pela utilização de um sistema sem que haja necessidade de uma instalação local em um servidor com *hardware* dedicado, já que o acesso ao serviço é feito totalmente *online* através de um computador, *tablet* ou até *smartphone*. Essa modalidade de licenciamento de *softwares* permite ao cliente ter acesso ao sistema de uma forma muito mais barata e rápida comparada aos modelos tradicionais de comercialização, isentando-o de problemas comuns ao manter *softwares* em estruturas próprias (DUBEY; WAGLE, 2007).

Ao mesmo tempo que esse tipo de solução facilita o acesso por parte dos clientes, gera uma maior concorrência entre as empresas, já que não existem mais barreiras geográficas na hora de escolher um fornecedor de *software*. Sendo assim, a otimização de custos operacionais passa a ser um fator crucial para a competitividade das empresas que fornecem esse tipo de solução no mercado.

Uma empresa de *softwares* para restaurantes situada na região de Porto Alegre, fornece acesso a um sistema para gerenciamento de restaurantes no modelo SaaS e possui um total de 610 clientes ativos atualmente. Ela conta com uma equipe de 35 profissionais, sendo que, 6 deles são responsáveis pelo suporte e 7 foram contratados para realização de treinamentos aos novos usuários. Ou seja, aproximadamente 37% da equipe contratada está atualmente alocada para estes fins e gera um custo de R\$ 27.000,00 em remuneração para a empresa estudada. O ticket médio referente ao sistema comercializado é de 200 reais ao mês e um novo usuário consome até 1 hora e 30 minutos de treinamento para iniciar a utilização.

Com uma taxa média de crescimento de 70 novos clientes ao mês, são necessárias até 105 horas/mês de treinamentos aos novos usuários. Atualmente, a empresa cobra contratualmente uma fidelização de 3 meses dos clientes, visando uma maior transparência e liberdade na contratação dos serviços e, devido aos dados citados anteriormente, as primeiras 6 mensalidades pagas são apenas para cobrir os custos de aquisição e implantação destes usuários. Isso faz com que seja crucial a satisfação dos clientes já que não existe muita caso os mesmos optem por não seguir utilizando o sistema após os três primeiros meses e o custo de implantação é superior ao retorno financeiro para a empresa neste período.

Existem diversos serviços com a função de auxiliar na adaptação de novos usuários em sistemas *web*, que visam otimizar os custos com suporte e possibilitar um crescimento mais rápido da base de clientes em um SaaS. Estes são chamados de Sistemas de *Onboarding* e utilizam componentes sobrepostos na tela para guiar o usuário nas ações que este deve realizar ao interagir com o sistema utilizado. Como exemplo para o presente trabalho será utilizado um serviço chamado *Appcues* (appcues.com) que conta com diversos serviços para o *User Onboarding*, entre eles, a criação de guias interativos, chamados na plataforma de “*Flows*”. Outros serviços que também oferecem soluções similares são: *Userlane* (userlane.com), *Intercom* (intercom.com) e *Userpilot* (userpilot.com).

Para que seja possível exibir estes guias no sistema desejado é necessário criar uma conta no *Appcues*, baixar uma ferramenta para criação dos guias (*Appcues Flow Builder*) e então cadastrar um novo *Flow* apontando para a URL do sistema em que se deseja aplicar a ferramenta. Após cadastrado o *Flow*, deve-se acessar a *Appcues Flow Builder* para iniciar o processo de configuração das etapas do guia e então definir em cada tela do sistema quais elementos irão aparecer, sua ordem, e suas características como: posicionamento, formato, cores, textos, ícones, entre outros (APPCUES, 2019). Essa abordagem de configuração acaba limitando a experiência do usuário, já que a ordem das ações acaba sendo pré-fixada e o mesmo não tem liberdade para executá-las de forma natural. Outro problema desse tipo de abordagem é que, uma vez que o sistema sofra alterações, parte do trabalho de configuração das telas poderá ter que ser refeito ou apresentará inconsistências aos usuários, gerando uma má experiência.

Para melhorar este cenário o presente trabalho buscará descobrir padrões de comportamento dos usuários já ativos do sistema, para assim prever possíveis ações a serem

tomadas em uma determinada tela por novos usuários. Com os modelos de predição treinados, será possível indicar caminhos comuns aos usuários e guiá-los de uma forma automática nas primeiras utilizações do sistema, além de melhorar a experiência de usuários já habituados ao sistema, facilitando o acesso aos recursos mais utilizados pelos mesmos sem a necessidade de uma configuração manual realizada por um técnico.

Como para este problema não existe um algoritmo facilmente descritível, dada a quantidade e complexidade das possíveis entradas (interações dos usuários), serão abordadas técnicas de aprendizado de máquina para encontrar uma solução viável. Para exemplificar o conceito de aprendizado de máquina pode-se considerar a tarefa de classificar *emails* entre *spam* ou legítimo. Para esta tarefa, sabe-se que a entrada será um *email*, formado basicamente por uma sequência de caracteres e que a saída será “sim” ou “não”, indicando se aquele *email* trata-se de *spam* ou não. Porém não existem algoritmos *hand crafted*, que definam regras estáticas, para transformar esta entrada nas saídas desejadas, considerando que este processo pode variar ao longo do tempo e de indivíduo para indivíduo. Por outro lado, pode-se facilmente compilar milhares de *emails* que já se sabe tratarem-se de *spam* e “aprender” o que determina tal classificação (ALPAYDIN, 2010). De uma forma sistemática, Mitchell (1997, p. 2) define aprendizado de máquina como: “Diz-se que um programa de computador aprende com a experiência E em relação a alguma classe de tarefas T e medida de desempenho P , se seu desempenho nas tarefas em T , medido por P , melhorar com a experiência E .”

O primeiro passo para a realização deste trabalho será a criação de um *plugin*, que será instalado em um sistema e realizará a captura das informações necessárias para a criação de um modelo preditivo. A cada interação dos usuários com o sistema, serão armazenadas em memória local informações sobre qual usuário realizou a ação, data e hora da interação, tela em que foi realizada e elemento que sofreu a ação (botões, campos de texto, etc), para então serem enviados para um servidor na nuvem em intervalos predeterminados. Após a coleta dos dados, o conteúdo resultante será manipulado para criação dos modelos preditivos através de aprendizado de máquina e então, os resultados obtidos serão apresentados em forma de gráficos e tabelas, contendo informações como a taxa de acerto do modelo e em que situações os acertos ocorreram.

Com isso será apresentado qual o percentual de acerto que os modelos podem ter ao determinar ações que os usuários devem tomar em uma tela e assim, possibilitar a

automatização do processo de criação de guias interativos. Caso a acurácia do modelo seja suficiente, será possível reduzir os custos com suporte e treinamento na empresa estudada e possibilitar um crescimento da base de clientes sem a necessidade de ampliação das equipes. Também serão abertas possibilidades para melhorias na *User Experience* do *software* com a utilização dos modelos para antecipar ações dos usuários e oferecer um sistema mais inteligente.

OBJETIVOS

Objetivo Geral:

O presente estudo tem por objetivo criar uma ferramenta que possibilite a captura de dados de interações dos usuários em um sistema *web*, permitindo induzir modelos preditivos para reconhecimento de padrões de comportamento destes usuários.

Objetivos Específicos:

- Identificar técnicas de aprendizado de máquina para resolução de problemas de reconhecimento de padrões;
- Criar uma ferramenta que possibilite a coleta e armazenamento do histórico de interações dos usuários em um sistema *web*;
- Extrair as informações necessárias para o treinamento dos modelos de predição.
- Utilizar aprendizado de máquina para reconhecimento dos padrões das interações capturadas;
- Avaliar os modelos treinados e determinar a acurácia das predições;
- Apresentar os relatórios de acurácia com os resultados obtidos.

METODOLOGIA

O presente trabalho caracteriza-se como uma pesquisa aplicada, pois trata-se do desenvolvimento de uma ferramenta para captura de uma massa de dados com o objetivo de definir um modelo preditivo baseado em aprendizado de máquina para reconhecimento de padrões em comportamento de usuários em um sistema *web*. A abordagem é classificada como quantitativa, apresentando resultados de acurácia dos experimentos realizados.

Será realizada uma pesquisa bibliográfica na área de aprendizado de máquina para embasamento teórico e entendimento das técnicas necessárias para resolução de problemas de reconhecimento de padrões e então, com base no estudo realizado, criada uma ferramenta para a coleta dos dados necessários para treinamento dos modelos preditivos. Em seguida, será feita instalação da ferramenta de coleta em um sistema existente e configurado um servidor para recebimento dos dados.

Após a captura dos dados será realizado um pré-processamento das informações obtidas para, em seguida, aplicar os algoritmos estudados e determinar taxas de assertividade nos modelos preditivos, apontando situações em que a técnica tem melhores resultados na área estudada.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Escrita do Anteprojeto	■	■		
Entrega do Anteprojeto		■		
Pesquisa bibliográfica		■	■	■
Criação da ferramenta para coleta de dados		■	■	
Captura dos dados de interações			■	■
Escrita do texto do TCC I	■	■	■	■
Revisão do texto do TCC I		■	■	■
Entrega do TCC I				■

Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
Pré-processamento da base de dados coletada	■			
Realização do treinamento dos modelos preditivos		■	■	
Escrita do texto do TCC II	■	■	■	■
Compilação dos resultados obtidos			■	■
Revisão do texto do TCC II	■	■	■	■
Entrega do TCC II				■

BIBLIOGRAFIA

ALPAYDIN, E. **Introduction to Machine Learning**. [s.l.] MIT Press, 2010.

APPCUES. **Appcues**. Disponível em: <<https://www.appcues.com/>>. Acesso em: 29 ago. 2019.

BUSINESS WIRE. **Global Software-as-a-Service (SaaS) Market (2018-2023)**. Disponível em: <<https://www.businesswire.com/news/home/20181114005369/en/Global-Software-as-a-Service-SaaS-Market-Outlook-2018-2023-Expected>>. Acesso em: 28 ago. 2019.

DUBEY, A.; WAGLE, D. Delivering software as a service. 2007.

MEIRELLES, F. S. **30ª Pesquisa Anual do Uso de TI nas Empresas, 2019**, 2019. Disponível em: <https://eaesp.fgv.br/sites/eaesp.fgv.br/files/pesti2019fgvciappt_2019.pdf>. Acesso em: 24 ago. 2019

MITCHELL, T. M. **Machine Learning**. [s.l.] McGraw-Hill Education, 1997.