



**EA006 / 2S2018**

Trabalho de Conclusão de Curso  
(Plano de Trabalho)

# **Aplicação de Árvores de Decisão para Previsão de Evasão de Clientes**

Aluna: Fernanda C. Queiroz

RA: 140618

**Orientador:**

Prof. Dr. Fernando A. C. Gomide

e-mail: [gomide@dca.fee.unicamp.br](mailto:gomide@dca.fee.unicamp.br)

# 1. Introdução

Plataformas multi-laterais são modelos de negócio que atuam facilitando as interações econômicas entre 2 ou mais grupos distintos de usuários [1]. Em muitos desses negócios, há a criação de efeitos de rede indiretos entre os diversos grupos que participam da plataforma. Um efeito de rede indireto ocorre quando o aumento da utilização do serviço pelos usuários de um grupo aumenta o valor percebido do serviço pelos usuários de outro grupo afiliado ao serviço. Um exemplo é o modelo de negócios de cartões de crédito. Consumidores querem cartões de bandeiras que são aceitas em vários estabelecimentos comerciais, enquanto que os estabelecimentos comerciais preferem bandeiras que são usadas por muitos consumidores. Sendo assim, o aumento do número de consumidores torna a plataforma mais valiosa para os estabelecimentos e vice-versa. De modo geral, a evasão de clientes (churn) é um problema enfrentado em várias indústrias pois impacta diretamente no faturamento através da perda da receita que poderia ser gerada por aqueles clientes. Em modelos de negócio que se apoiam em efeitos de rede indireto, torna-se ainda mais importante prever e evitar que os clientes abandonem o serviço pois isso implicaria não somente na perda direta de receita, mas também na perda de valor percebido pelos demais usuários do serviço. Um cliente pode voluntariamente decidir abandonar uma plataforma de serviços por razões tecnológicas (clientes querem uma tecnologia mais moderna ou melhor), economica (sensibilidade ao preço), qualidade do serviço, fatores sociais, psicológicos e de conveniência [2]. Prever que um cliente irá abandonar o serviço antes que ele o faça representa uma grande economia porque o custo de aquisição de um cliente costuma ser muito maior do que o custo de rete-lo. Uma das técnicas mais tradicionais para a criação de modelos preditivos do comportamento desses clientes é a criação de árvores de decisão. Árvores de decisão são formas de classificar o estado de uma variável de saída (neste caso, definir se o cliente possui ou não risco de evasão) a partir da verificação do valor de múltiplas variáveis de entrada (como por exemplo, a frequência de pedidos, o valor gasto e a data da última compra realizada por este cliente) [3]. Uma das principais vantagens deste tipo de método de classificação é que, além de facilitar o processo de tomada de decisão, ele é facilmente interpretável, contribuindo para identificar os fatores causadores da evasão.

## 2. Objetivos

O presente trabalho consiste em endereçar o problema de evasão de clientes em organizações que produzem valor majoritariamente através de efeitos de rede indiretos. Construindo um modelo de árvore de decisão, pretende-se encontrar um método de classificação que concilie a interpretabilidade dos fatores causadores de evasão com uma boa precisão na identificação desses clientes.

## 3. Metodologia

A proposta inicial desse trabalho é estudar algoritmos de classificação através de árvores de decisão enquanto obtem-se permissão para a utilização de dados privados de uma empresa do perfil escolhido. Depois disso os dados seriam coletados e tratados, anonimizando informações confidenciais, removendo inconsistências e deixando-os em uma estrutura que facilite a modelagem. A construção do modelo em si será feita selecionando os atributos mais relevantes da base de dados e ajustando os parâmetros do algoritmo de otimização até se obter uma boa precisão em prever os clientes que irão abandonar o serviço.

## 4. Cronograma

### Agosto

- Estudo dos algoritmos de classificação
- Obtenção dos dados

### Setembro

- Anonimização dos dados
- Remoção de inconsistências
- Preparar a base para a modelagem

### Outubro

- Implementação do modelo de decisão em árvore
- Múltiplas iterações para otimizar os parâmetros do modelo

### Novembro

- Confecção do relatório final
- Elaboração do poster

## 5. Referências

- [1] HAGIU, A; WRIGHT, J. *Multi-Sided Platforms*. International Journal of Industrial Organization, Vol. 43, 2015.
- [2] ESSAM, S; YEHA, H; AYMAN, K; MONA, N. *A Proposed Churn Prediction Model*. International Journal of Engineering Research and Applications. Vol. 2, Issue 4, June-July 2012, pp.693-697
- [3] PROVOST, F; FAWCETT, T. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. 1. ed. O'Really Media, Inc., Sebastopol. 2013