

Aplicação de Mineração de Dados no gerenciamento do Churn em Startup do segmento SaaS

Antônio Roque Falcão Júnior, Rafael Ballottin Martins

Laboratório de Inteligência Aplicada – Escola do Mar Ciência e Tecnologia –
Universidade do Vale do Itajaí (UNIVALI)
Rua Uruguai, 458 - 88302-202 – Itajaí- SC – Brasil

arfj@edu.univali.br, ballottin@univali.br

Abstract. *Expert's experience and intuition are no longer sufficient to understand the rapid changes in the profile of consumers, as well as the main motivations of Churn, in this scenario, companies and organizations are induced to strengthen the decision-making process. The search for knowledge through alternative technologies highlights Data Mining as a means to extract knowledge in data sets, which in most cases has its use limited to trivial and little exploratory activities. Through this research was performed, with the help of Data Mining process, the identification of standards and extraction of knowledge, used to answer research questions.*

Resumo. *A experiência e a intuição do especialista já não são suficientes para compreender as rápidas mudanças no perfil dos consumidores, assim como as principais motivações do Churn, neste cenário, empresas e organizações são induzidas a fortalecer o processo de tomada de decisão. A busca por conhecimento através de tecnologias alternativas destaca a Mineração de Dados como um meio de extrair conhecimento em conjuntos de dados, que na maioria dos casos tem seu uso limitado a atividades triviais e pouco exploratórias. Através desta pesquisa foi realizada, com o auxílio do processo de Mineração de Dados, a identificação de padrões e extração de conhecimento, utilizados para responder questões de pesquisa.*

1. Introdução

Segmentos comerciais de crescimento acelerado, nascidos na “era da informação”, considerados Startups, se desenvolvem em ambiente competitivo e de rápidas mudanças. A informação é fundamental para sobrevivência deste tipo de organização, não apenas a informação/dado em seu estado natural, estático ou físico, mas sim, sua forma de gestão e entendimento [SEBRAE e CNPQ 2014].

O *Churn*, é referente a quantidade de clientes ou assinantes que encerram o relacionamento com os seus serviços, produtos ou fornecedores, definitivamente ou durante um determinado período de tempo. Esses clientes, provém as altas nos índices do Churn [Chen et al. 2015].

De acordo com dados da *Pacific Crest* pelo menos 30% de todo conglomerado tecnológico do ramo possui números inaceitáveis para este indicador. Valores saudáveis giram em torno de 5 a 7% ao ano. Identificar a origem e controlar o *Churn* é uma questão de sobrevivência [Mehta et al. 2016].

O modelo de negócio SaaS (*Software as a Service*) tem por base o software como serviço, comercializado no modo assinatura e não mais com a tradicional venda de licenças vitalícias. A mudança não se caracteriza apenas na comercialização, mas também pelo novo perfil de relacionamento “Cliente x Empresa” adotado pelo fornecedor, neste conceito, o próprio distribuidor passar a atuar para garantir a permanência do usuário na plataforma [Carraro e Chong 2006].

Reter para crescer, não é apenas um estilo de gestão, mas uma necessidade para o mercado de serviços. Buscando atender tal demanda, sugere-se o uso de técnicas de *Onboarding*, que se caracterizam principalmente pelo acompanhamento do cliente

recém-chegado durante o período de adaptação. Um dos principais objetivos é evitar que o novo usuário tenha uma primeira má impressão e abandone o produto antes que o mesmo tenha oportunidade de agregar valor ao seu negócio [Hulick 2014].

A interação entre usuários e serviços, somados a evolução das tecnologias de armazenamento, assim como dos sistemas gerenciadores de banco de dados, resultam em grande volume de dados armazenados pelas empresas, em sua grande maioria utilizados apenas em atividades triviais e pouco exploratórias [Provost e Fawcett 2013].

Diante deste cenário, empresas e organizações são induzidas a fortalecer o processo de tomadas de decisão. A busca por conhecimento através de tecnologias alternativas destaca a Mineração de Dados (MD) como meio para extração de conhecimento em conjuntos de dados.

Mineração de Dados é a etapa de análise do processo de descoberta de conhecimento. A Mineração de Dados pode ser dividida em algumas etapas básicas como: exploração, construção de modelo, definição de padrão, validação e verificação. Utiliza técnicas de clusterização, classificação, regras de associação, mineração de textos, IA (Inteligência Artificial), dentre outras. Além de viabilizar recuperação de informações, reconhecimento de padrões e estatísticas [Goldschmidt e Passos 2015].

O KDD (*Knowledge Discovery in Databases*) é um processo não trivial de identificar dados válidos, de contexto não superficial, novos e potencialmente úteis em banco de dados. O processo KDD pode ser visto como uma atividade multidisciplinar que engloba técnicas além do âmbito geral, ou de qualquer disciplina específica, como Aprendizado de Máquina, Mineração de Dados, dentre outras. O KDD coloca uma ênfase especial para padrões compreensíveis que podem ser interpretados como conhecimento ou informações úteis ou interessantes [Fayyad et al. 1980].

Com auxílio do processo de *Knowledge Discovery in Databases* (KDD) como meio de integração dos dados, extração e análise de informações, em conjunto às tarefas de Mineração de Dados, buscou-se gerar conhecimento de negócio sobre o desempenho comercial de uma startup do segmento SasS, com ênfase ao gerenciamento do *Churn*.

2. Conjunto de dados

Os dados utilizados foram gerados a partir de 4 sistemas de diferentes setores e com diferentes propósitos, dentre eles: O sistema do setor Administrativo que serve como plataforma para gestão de clientes e licenças; Do setor Comercial que é utilizado para controle do fluxo de venda e etapas do Onboarding; Do setor de Marketing que é a solução utilizada na captação de novos clientes ou parceiros; e do setor de produção que permitiu coletar maiores detalhes a respeito do uso deste serviço.

Além das fontes de dados mencionadas, foram disponibilizadas 2 planilhas, contendo registro da performance de duas das principais KPIs da empresa.

A plataforma escolhida para integração dos dados foi o MS-SQL Server, devido a base principal, nativa desta plataforma, ser o conjunto a apresentar o maior volume de dados, cerca de 80 gigabytes, dificultando assim sua exportação para outra ferramenta.

3. Modelagem dos dados

Inicialmente com a utilização de um *Data Warehouse* (DW), buscou-se confrontar as hipóteses de que a completude do processo de *Onboarding* e a frequência de acesso à plataforma influenciam positivamente a taxa do *Churn*, assim como procurou-se identificar fatores que levassem à perda dos clientes.

O *Churn* foi tratado como rótulo a ser descoberto, de forma que os demais atributos se tornaram fatores que poderiam levar ao fenômeno. Nesse contexto, optou-se por utilizar algoritmos classificadores, que gerassem regras ou árvores de decisão como

saída e que fossem capazes de lidar com rótulos. De acordo com a análise realizada por Carvalho (2010), a autora obteve a melhor performance de acertos com o algoritmo ID3 através de benchmark contendo o J48 e o *SimpleCart*.

Com o propósito de responder perguntas de pesquisa a respeito das características ou motivações para o *Churn*, e da influência de determinadas KPIs na retenção de clientes, buscou-se a relação entre o fenômeno e os demais atributos no perfil dos clientes. Nesse caso, as duas abordagens foram tratadas de forma a apresentar um contexto maior a análise dos dados. Segundo Fayyad et al. (1980), em aplicações de MD, as regras de associação funcionam de modo a identificar as operações complexas. Os conjuntos de artigos frequentes devem ser gerados anteriormente para se obter essas regras de associação.

Para esta pesquisa, dois dos algoritmos mais comuns foram escolhidos, para este tipo de ação: *Apriori* e *FP-Growth*. O *Apriori* gera os conjuntos frequentes e as regras de associação, já o *FP-Growth* gera apenas conjuntos frequentes, que são então utilizados pelo operador *Create Association Rules* para gerar as regras de associação. Para realização das tarefas foi utilizada a ferramenta RapidMiner [Hofmann e Klinkenberg 2013] que possui suporte para todas as etapas do KDD.

4. Resultado da Classificação

A partir da execução algoritmo ID3 a árvore foi configurada para obter um ganho mínimo de informação de 0.1, acurácia como critério e 4 como tamanho mínimo para dividir. Os valores foram determinados após testes indicarem que a árvore não seria montada se o ganho de informação fosse inferior a 0.1. O conjunto de dados estudado possui 9 atributos, onde um deles é utilizado como rótulo. O fator de confiança para este algoritmo não é parametrizável pela ferramenta, onde para visualiza-lo, junto aos resultados, é necessário replicar o modelo através do operador *Apply Model* e medir os resultados com operador *Performance*.

Uma das saídas geradas pela classificação é a representação gráfica das regras, em forma de árvore. 60 regras, foram geradas ao final do processo, com confiança mínima de 0.443 e máxima de 0.854, conforme Tabela 1.

Tabela 1. Regras de classificação

Nº	Regra
1	if KPI_ATM = AbaixoDoIndice then SIM
2 - 5	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_1000 and MesesDeUso in (Entre_3_e_6, Entre_7_e_10, Entre_11_e_18, MaiorQue_18) then SIM
6	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_1000 and MesesDeUso = MenorQue_3 and Conversão = Busca Orgânica and EnquadramentoFiscal = MEI then SIM
7 - 8	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_1000 and MesesDeUso = MenorQue_3 and Conversão = Busca Orgânica and EnquadramentoFiscal in (Outros, Simples Nacional) then NÃO
9	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_1000 and MesesDeUso = MenorQue_3 and Conversão = Busca Paga then SIM
10, 27	if KPI_ATM = DentroDoIndice and QtdMediaNFE in (Aprox_3000, Aprox_5000) then SIM
11	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = ME then NAO
12	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = MEI and Conversão = Busca Orgânica then NAO
13	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = MEI and Conversão = Busca Paga then SIM
14	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Outros and MesesDeUso = Entre_11_e_18 then SIM
15, 17	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Outros and MesesDeUso = Entre_3_e_6 and FrequenciaAcesso in (Diário, Semanal) then NÃO
18 -19	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Outros and MesesDeUso in (Entre_7_e_10, MaiorQue_18) then SIM

20, 22	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Outros and MesesDeUso = MenorQue_3 and FrequenciaAcesso in (Diário, Quinzenal) then NÃO
23-24	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Simples Nacional and MesesDeUso in (Entre_11_e_18, Entre_3_e_6) then NÃO
25	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Simples Nacional and MesesDeUso = MenorQue_3 and Conversão = Busca Orgânica then SIM
26	if KPI_ATM = DentroDoIndice and QtdMediaNFE = Aprox_500 and EnquadramentoFiscal = Simples Nacional and MesesDeUso = MenorQue_3 and Conversão = Busca Paga then NÃO
28	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Diário then SIM
29	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Mensal and EnquadramentoFiscal = Outros then SIM
30	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Mensal and EnquadramentoFiscal = Simples Nacional then NÃO
31	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Quinzenal and Onboarding = Concluso and EnquadramentoFiscal = Outros then SIM
32	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Quinzenal and Onboarding = Concluso and EnquadramentoFiscal = Simples Nacional then NÃO
33	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Quinzenal and Onboarding = Parcial then SIM
34	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_11_e_18 and FrequenciaAcesso = Semanal then SIM
35	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_3_e_6 and Onboarding = Concluso then NÃO
36	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_3_e_6 and Onboarding = Parcial then SIM
37	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = Entre_7_e_10 then SIM
38	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Lucro Real ou Presumido then SIM
39	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Outros and FrequenciaAcesso = Diário then NÃO
40 - 41	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Outros and FrequenciaAcesso = Quinzenal and QtdFilial = Entre_3_e_6 then SIM
42	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Outros and FrequenciaAcesso = Semanal then NÃO
43	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Simples Nacional and FrequenciaAcesso = Quinzenal then SIM
44	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MaiorQue_18 and EnquadramentoFiscal = Simples Nacional and FrequenciaAcesso = Semanal then NÃO
45	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Lucro Real ou Presumido and FrequenciaAcesso = Semanal then NÃO
46, 48	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = ME and FrequenciaAcesso in (Diário, Semanal) then NÃO
49	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = MEI then NÃO
50-51	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Outros and Conversão = Busca Orgânica and Onboarding = Concluso and FrequenciaAcesso in (Mensa, Quinzenal) then SIM
53,55	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Outros and Conversão = Busca Orgânica and Onboarding = Parcial and FrequenciaAcesso in (Diário, Semanal) then NÃO

54,56	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Outros and Conversão = Busca Orgânica and Onboarding = Parcial and FrequenciaAcesso in (Mensal, Quinzenal) then SIM
57	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Outros and Conversão = Busca Paga and Onboarding = Concluso and then NÃO
58	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Outros and Conversão = Busca Paga and Onboarding = Parcial and then SIM
59	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Simples Nacional and FrequenciaAcesso = Diário then NÃO
60	if KPI_ATM = DentroDoIndice and QtdMediaNFE = MenorQue_100 and MesesDeUso = MenorQue_3 and EnquadramentoFiscal = Simples Nacional and FrequenciaAcesso = Mensal then SIM

Na primeira ramificação da árvore, o algoritmo assumiu como o nó inicial o atributo KPI_ATM representando a faixa do índice da KPI de satisfação do atendimento, dividindo as regras em dois grupos: Abaixo do índice e Dentro do índice. Para a faixa Dentro do índice o algoritmo separou pela quantidade média de Notas Fiscais emitidas, a partir de então, novas divisões ocorreram, sobre do tempo de uso do sistema e enquadramento fiscal.

De acordo com os resultados, para 17,71% dos casos analisados, o *Churn* ocorreu em períodos onde a satisfação dos clientes com o atendimento esteve abaixo do índice da KPI. Nos demais casos, a quantidade média de Notas Fiscais emitidas por mês foi determinante ao *Churn* para 8,77% dos casos analisados, na qual as emissões superaram 3.000 notas.

Mesmo para os 12,06% dos casos onde clientes emitiram em torno de 1000 notas, a desistência ocorreu em todos os casos em que o período de uso foi maior que três meses, percentualmente 3,08%. O número de clientes perdidos diminui à medida em que o número de notas fiscais emitidas diminui e se aproxima de 500.

Empresas optantes do Simples Nacional, ou de enquadramento fiscal ME e MEI, perfis característicos do micro e pequeno varejo do Brasil, apesar de representarem 67% de todo conjunto, possuem recorrência de *Churn* para apenas 18,76% dos casos, que em sua maioria é acompanhado da frequência de acesso mensal.

O padrão para o *Churn* determinado pela frequência de acesso ocorre em 47,2% dos casos para Mensal e 28,12% para Quinzenal. Contudo, as frequências Diário e Semanal juntas representam apenas 24,68% das ocorrências. O padrão foi observado em todas as sub árvores geradas a com o número de notas menor ou igual a 500.

Outro importante padrão observado, diz respeito à origem de conversão/captação dos clientes via mídias digitais. Para este fator a qualidade é medida pela permanência do cliente na plataforma com base na sua origem de captação, se através de busca Paga ou Orgânica.

A representatividade da captação Orgânica na classificação é de 63,7% contra 36,3% de captação Paga. Contudo, e no que diz respeito ao *Churn*, 71,04% dos clientes que foram captados via buscas pagas abandonam o produto, contra 23,47% das captações de cliente que ocorreram via buscas orgânicas.

Dentre os clientes que passaram pelo processo de *Onboarding*, 67,6% não superaram/completaram todas as etapas. Para os 32,4% dos registros da classificação que atingiram todos os níveis do processo de acompanhamento, apenas 7,2% vieram a se tornar *Churn*, contra 28,2% dos 67,6% que não superaram todas as etapas do processo.

5. Resultado da Associação

A análise realizada pelo algoritmo considera as características de determinado registro/cliente como itens a serem associados. As regras de associação foram criadas considerando um suporte mínimo (Sup) de 0.122 e uma confiança (Conf) superior a 0.9. Esses números foram definidos com base em testes e análise dos resultados na perspectiva dos gestores da empresa.

O desafio neste processo, consistiu da necessidade de gerar regras que atendessem a expectativa de associação dentre pelo menos 6 dos 9 atributos selecionados, e que possuíssem o *Churn* como conclusão.

A Tabela 2 apresenta as principais regras que atingiram as configurações determinadas e estão ordenadas pela quantidade de itens associados.

Tabela 2. Regras de associação

Nº	Premissa	Conclusão	Sup	Conf
1	KPI_ATM = AbaixoDoIndice	Churn = SIM	0.177	1.0
2	Onboarding = Parcial, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.177	1.0
3	Conversão = Busca Orgânica, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.128	1.0
4	QtdFilial = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.173	1.0
5	FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.177	1.0
6	MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.174	1.0
7	Onboarding = Parcial, Conversão = Busca Orgânica, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.128	1.0
8	Onboarding = Parcial, QtdFilial = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.173	1.0
9	Onboarding = Parcial, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.177	1.0
10	Onboarding = Parcial, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.174	1.0
11	Conversão = Busca Orgânica, QtdFilial = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.124	1.0
12	Conversão = Busca Orgânica, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.128	1.0
13	Conversão = Busca Orgânica, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.126	1.0
14	QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.173	1.0
15	QtdFilial = MenorQue_3, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.170	1.0
16	FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.174	1.0
17	Onboarding = Parcial, Conversão = Busca Orgânica, QtdFilial = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.124	1.0
18	Onboarding = Parcial, Conversão = Busca Orgânica, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.128	1.0
19	Onboarding = Parcial, Conversão = Busca Orgânica, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.126	1.0
20	Onboarding = Parcial, QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.173	1.0
21	Onboarding = Parcial, QtdFilial = MenorQue_3, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.170	1.0
22	Onboarding = Parcial, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.174	1.0
23	Conversão = Busca Orgânica, QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.124	1.0
24	Conversão = Busca Orgânica, QtdFilial = MenorQue_3, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.121	1.0
25	Conversão = Busca Orgânica, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.126	1.0
26	QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.170	1.0
27	Onboarding = Parcial, Conversão = Busca Orgânica, QtdFilial =	Churn = SIM	0.124	1.0

Nº	Premissa	Conclusão	Sup	Conf
	MenorQue_3, FrequenciaAcesso = Mensal, KPI_ATM = AbaixoDoIndice			
28	Onboarding = Parcial, Conversão = Busca Orgânica, QtdFilial = MenorQue_3, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.121	1.0
29	Onboarding = Parcial, Conversão = Busca Orgânica, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.126	1.0
30	Onboarding = Parcial, QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.170	1.0
31	Conversão = Busca Orgânica, QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.121	1.0
32	Onboarding = Parcial, Conversão = Busca Orgânica, QtdFilial = MenorQue_3, FrequenciaAcesso = Mensal, MesesDeUso = MenorQue_3, KPI_ATM = AbaixoDoIndice	Churn = SIM	0.123	1.0

A regra de número 32 representa a associação de sete dos nove itens possíveis, com suporte de 0.123 e confiança de 100%, que por sua vez, melhor atende ao padrão buscado.

Dois atributos não foram passíveis de associação, que correspondiam ao enquadramento fiscal e quantidade média de notas fiscais emitidas por mês. Foi realizado testes com o suporte em 0.05 e em 0.01, ambos com confiança de 0.7. Mesmo com essas configurações, não foi possível observar tais itens em associações com o mínimo de quatro elementos.

6. Análise dos resultados da Classificação

A empresa pesquisada, ao longo dos seus cinco anos de atividade acumulou 52,7% de *Churn*, a uma média de 10,54% ao ano. Em especial para ano de 2017, em média até o mês de setembro, onde esse número é ainda maior, 12,07% ao mês. Cerca de 5,07% acima do que recomendado pelo especialista no assunto, Murphy et al. (2016).

Uma das principais motivações para a perda de clientes, de acordo com o entendimento dos analistas do negócio, está ligado à qualidade do que é captado. Por ser uma solução voltada para o micro e pequeno varejista, o produto não consegue atender demandas específicas de empresas que atuam em regimes tributários maiores, a exemplo do Lucro Real e Presumido, que representa 23,5% das captações. Filtrar melhor a entrada de potenciais clientes foi uma das alternativas discutidas. As regras 2,3,4 e 5 comprovam que clientes que emitem mil ou mais notas fiscais por mês, volume atípico para empresas micro e pequenas, tornam-se *Churn*, percentualmente 68% deste volume. Para os 32% restantes apenas 6% está enquadrado como Outros, os demais que emitem próximo a esse volume de notas o enquadramento fiscal é o Simples Nacional, MEI e ME, que por sua vez, permanecem na plataforma.

Ao confrontar a hipótese 1 desta pesquisa, nos deparamos com a afirmação de que superar as etapas do *Onboarding* reduz as chances do cliente se tornar *Churn*. De acordo com as regras de classificação 35 e 57, os clientes que completaram o processo não se tornaram *Churn* mesmo após o sexto mês de uso. De forma complementar as regras 33, 36, e 58 reafirmam que a incidência de abandono é mais comum para os casos onde o acompanhamento tenha sido encerrado antes da conclusão. Para regra 58 o mesmo ocorre antes de ser completado o terceiro mês de uso. Somadas, as cinco regras envolvem 53,2% dos casos analisados. Não foram considerados na soma os casos que passam por tais atributos, a exemplo das regras 31, 32, 50, 51, 52, 53, 54, 55 e 56, passíveis de agregar volume ao percentual analisado. Diante da análise dos resultados apresentados, os gestores do negócio concluíram que a hipótese é verdadeira.

Para testar a hipótese 2 desta pesquisa, que afirma que a frequência de acesso ao produto possui influência sobre o índice do *Churn*. Foi necessário analisar todas as regras que resultassem no atributo “FrequenciaAcesso” e calcular tais resultados. De acordo com as regras 15, 20, 39, 46, 53 e 59, os clientes que acessam o sistema diariamente não são propícios ao *Churn*. Mesmo para os casos onde a frequência de acesso é semanal, a exemplo das regras 17, 42, 44, 48, 52 e 56, ainda assim, o fenômeno não ocorre. Dando seguimento a análise, nos deparamos com a frequência quinzenal, e junto a ela a regra 22 reafirmando a não ocorrência do *Churn*. Contudo, e de contrapartida, as regras 43, 51 e 55, para a mesma frequência, afirmam que o *Churn* ocorre. Ao incluirmos na análise a frequência mensal, averiguou-se que para as regras 16, 21, 47, 50, 54 e 60, que por sua vez representam 100% das frequências mensais, o *Churn* acontece. De acordo com os resultados, para 7% dos casos onde a frequência de acesso é diária ou semana o *Churn* acontece, e que para 93% destes não ocorre. Por outro lado, 10% das frequências quinzenais e mensais não possui *Churn*, onde seus 90% confirmam como verdadeira a hipótese 2, com base nos dados analisados e entendimento dos representantes da empresa.

A regra de classificação 1 é objetiva, indica que o *Churn* ocorreu para 17,71% dos casos analisados. Estes usuários possuíam como característica, o fato de terem sido clientes em períodos onde o índice de satisfação do atendimento estava abaixo do esperado para a KPI. Acreditava-se haver a relação entre a qualidade do atendimento e a retenção dos clientes, porém os dados demonstram que isso pode ser ainda mais relevante. A classificação não foi capaz de gerar qualquer outra regra contendo o atributo KPI_ATM como resultado, pois toda a ramificação direta possui o valor acima do índice. Apesar de única, neste contexto, a regra satisfez o questionamento por demonstrar que para todos os casos onde o índice de satisfação esteve abaixo do esperado, o *Churn* veio a ocorrer. Segundo os gestores, no que diz respeito a KPI de completude do *Onboarding*, a mesma já havia sido respondida em meio a análise da hipótese 1, considerando assim como respondida a uma das perguntas da pesquisa, de forma afirmativa para a existência de KPIs que podem influenciar positivamente a retenção dos clientes.

7. Análise dos resultados da Associação

O objetivo desta tarefa, foi prover um modelo capaz de gerar dados com potencial para responder à pergunta principal desta pesquisa, que se refere às principais características do *Churn*. Diante deste contexto, decidiu-se por associar os principais atributos a fim de identificar a maior frequência de características (atributos) quanto fosse possível. A regra 32 foi capaz de obter como resultado o *Churn* associando outros 6 atributos\características do fenômeno, para o modelo de negócio em estudo.

Considerando a objetividade do processo em associar as frequências, vimos a possibilidade de reforçar algumas questões anteriormente comprovadas com a classificação, em relação ao *Churn*, a exemplo da frequência de acesso mensal, *Onboarding* parcial e o índice de satisfação do atendimento abaixo da KPI serem fatores que influenciam a perda de clientes.

8. Conclusões

O desenvolvimento do trabalho implicou um estudo sobre como a empresa pesquisada funcionava, com foco no gerenciamento do *Churn*, e compreensão das principais motivações para perda de cliente no segmento SaaS.

A utilização da metodologia KDD forneceu um importante guia em como trabalhar em todas as fases do processo de mineração de dados, que foi imprescindível para a conclusão desse trabalho, uma vez que a quantidade de dados e atributos tornava uma análise manual inviável.

Foi de extrema importância a escolha da ferramenta a ser utilizada durante todas as partes do processo. O Rapidminer se fez importante para a pesquisa, uma vez que foi capaz de associar-se ao banco de dados para facilitar o processo de extração, dispor de recursos fundamentais e práticos para transformação dos dados, além de contar com uma usabilidade diferenciada em relação a outras ferramentas.

Junto aos gestores do negócio foi possível identificar que dentre os maiores problemas, destaca-se a falta de informação clara a respeito das principais motivações para o abandono do serviço/produto por parte dos clientes. Além de contar com o auxílio para compreensão adequada das hipóteses e questionamentos que fossem essenciais para a organização.

O algoritmo *FP-Growth* e *Create Association Rules* foram utilizados para buscar características padrões entre o *Churn*. Em termos de performance, a técnica utilizada respondeu de forma satisfatória à expectativa dos *stakeholders*, associando 7 atributos dos 9 possíveis. Por experiência os gestores acreditavam que a frequência de acesso mensal, a completude do *Onboarding* e a qualidade do atendimento influenciavam a retenção dos clientes, e elas se confirmaram nas regras geradas.

O algoritmo ID3 foi utilizado para criar uma árvore de decisões e buscar pelos fatores que levam ao *Churn*. A árvore foi capaz de identificar padrões para o *Churn* a partir do enquadramento fiscal, índice de satisfação do atendimento, completude do processo de acompanhamento, número de notas fiscais emitidas, frequência de acesso e captação dos clientes.

Ficou claro que clientes captados via busca paga, possuem maior tendência a abandonar o serviço além de representarem um custo maior para a empresa. Diante de tal informação, os diretores estudam investir maior atenção às captações orgânicas, a fim de reduzir custo e o *Churn*. Além deste fator, também ficou evidente que clientes maiores, de regimes tributários incomuns ao micro e pequeno varejo possuem grande incidência de *Churn*, e diante desta informação, os gestores estudam alternativas para comunicar com maior clareza aos possíveis clientes o posicionamento do produto/serviço no mercado, assim como refinar o filtro das captações de clientes, para que sejam melhor direcionadas ao público alvo (Micro e pequenos varejistas).

No que diz respeito ao processo de acompanhamento (*Onboarding*), os resultados obtidos serviram de aporte para a decisão da permanência do mesmo no fluxo de aquisição de novos clientes. Ante ao desenvolvimento desta pesquisa, haviam dúvidas quanto a sua eficácia e influência sobre a retenção de clientes na empresa.

Em relação a influência da frequência de acesso sobre a permanência dos clientes na plataforma, foi demandado aos gerentes de produto, pelos gestores, que estudassem funcionalidades e inovações que permitam um acompanhamento mais detalhado para os clientes em relação a seus empreendimentos, trazendo assim a gestão dos negócios ainda mais para o âmbito do produto, com principal foco em suas necessidades diárias.

Outro importante ponto de atenção abordado pela pesquisa, foi a forte influência da qualidade/satisfação do atendimento sobre a desistência de clientes. De acordo com os gestores tal KPI passará a ter seu monitoramento reduzido de mensal para quinzenal.

Referências

- Carraro, G.; Chong, F. *Software as a Service (SaaS): Enterprise Perspective*. 2015
- Carvalho, R. A. S. *Data mining no contexto de Customer Relationship management em uma franquia Coca-Cola Company*, 2010. Universidade Federal de Pernambuco.
- Chen, K.; Hu, Y.-H.; Hsieh B.; Arthur Y.-C. *Predicting customer churn from valuable B2B customers in the logistics industry: a case study*. In: *Information Systems and e-Business Management*, v. 13, n. 3, p. 475–494, 2015. Springer Berlin Heidelberg.

- Fayyad, U.; Smyth, P.; Piatetsky-Shapiro, G. From Data Mining to Knowledge Discovery in Databases. 1o ed. Providence, Rhode Island: American Association for Artificial Intelligence, 1980.
- Goldschmidt, R.; Passos, E. Datamining. Conceitos, Técnicas, Algoritmos, Orientações e Aplicações. 2o ed. Rio de Janeiro: CAMPUS, 2015.
- Hofmann, M., Klinkenberg, R. RapidMiner: Data Mining Use Cases and Business Analytics Applications, CRC Press, 1st edition, 2013.
- Hulick, S. The Elements of User Onboarding by Samuel Hulick — Reviews, Discussion, Book clubs, Lists. 1o ed. New York, 2014.
- Mehta, N., Steinman, D., Murphy, L. Customer success: how innovative companies are reducing churn and growing recurring revenue. 1 ed. New York: Wiley, 2016.
- Provost, F.; Fawcett, T. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. 1o ed. Sebastopol: O'Reilly Media, 2013.
- SEBRAE, CNPQ. Comércio e Serviços. Cadernos de inovação em pequenos negócios, 2014.