

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 – Centro de Ciências Tecnológicas da Terra e do Mar – 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.*

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 and CNPQ 2014].

O *Churn*, também conhecido como taxa de abandono, desistência ou cancelamento, é o principal responsável pelo insucesso da maioria das Startups. De acordo com dados da *Pacific Crest* pelo menos 30% de todo conglomerado tecnológico do ramo possui números inaceitáveis para este indicador [Mehta et al. 2016].

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 como meio para extração de conhecimento em conjuntos de dados, que em grande parte dos casos tem seu uso limitado a atividades triviais e pouco exploratórias.

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 da empresa, com ênfase ao gerenciamento do *Churn*.

2. Solução proposta

Nesta pesquisa objetivou-se aplicar técnicas de Mineração de Dados junto aos princípios do KDD, a fim de identificar padrões e extrair conhecimento em base de dados em uma Startup catarinense que fornece solução de modelo SaaS (*Software as a Service*) para

automação de micro e pequenas empresas do varejo. Com o propósito de identificar as principais motivações do *Churn*, mapear características de perfis dos usuários que desistem do produto, assim como validar a influência do *Onboarding* na retenção de clientes.

3. Projeto

Para um melhor entendimento do modelo de negócio, assim como suporte à formulação das perguntas de pesquisa, o auxílio dos analistas do negócio, tais como diretores de marketing, vendas e administrativo, se fez necessário.

Dois bancos de dados relacionais, foram disponibilizados para o processo de descoberta. Além dos registros gerados pelas ferramentas de gestão do time de marketing e da equipe vendas, contendo dados do relacionamento com o cliente quanto às etapas de captação, negociação e acompanhamento. De forma complementar, foi concedido acesso às planilhas de controle de dois dos principais indicadores de performance da empresa, referentes ao índice de satisfação do atendimento e completude do processo de acompanhamento.

Para a devida familiarização com o conjunto de dados alvo deste projeto, amostras foram submetidas a análise de caráter exploratório, observou-se a qualidade dos dados, assim como o potencial das informações ali contidas, norteando quanto aos potenciais atributos a serem utilizados.

Após pesquisas, foi selecionada a ferramenta para mineração dos dados utilizando como referência o trabalho realizado por Augusto Junior (2015), em que o autor aborda de forma sistemática diversas ferramentas para análise e mineração de dados com o propósito de “minimizar um viés na escolha”. O autor ainda descreve uma série de requisitos importantes para as ferramentas, compatíveis com esta pesquisa, entre elas, a licença de uso livre para academia, fornecer mais de um algoritmo para tarefas de Classificação, Associação e Clusterização, não ser especializada em um só tipo de algoritmo e ter uma versão relativamente recente. Além disso foram analisados a usabilidade, desempenho e aceitação da comunidade. Com base nos fundamentos apresentados da respectiva pesquisa, para o desenvolvimento desta, optou-se pela ferramenta Rapidminer, por sua interessante abrangência e viabilidade de uso dentro do escopo que este trabalho busca.

Uma vez composto o *Data Warehouse* (DW) através da integração dos dados dos diferentes conjuntos, foram definidas as tarefas e técnicas a serem utilizadas na etapa de mineração de dados, para extração de conhecimento através dos processos de descoberta supervisionada e não-supervisionada.

As tarefas de classificação e associação foram definidas para utilização no processo de descoberta de conhecimento, servindo de aporte para resposta aos questionamentos de pesquisa.

4. Considerações finais

A utilização da metodologia KDD forneceu um importante guia de como trabalhar em todas as fases do processo de mineração de dados, imprescindível para a realização desse trabalho.

Para cumprimento dos objetivos necessários até este momento da pesquisa, testes preliminares foram realizados com a intenção de identificar as tarefas, resultando na definição da tarefa de classificação para o Churn e associação com base nas características da relação de registros/clientes, os algoritmos *Decision Tree* e ID3 para classificação, o algoritmo Apriori e a combinação *FP-Growth* e *Create Association Rules* para a tarefa de associação, ambas com base na característica dos dados/problema em relação as características do algoritmo.

Por fim, também foram definidas as melhores configurações para que os objetivos sejam alcançados. Para a classificação, os algoritmos foram configurados para obter um ganho mínimo de informação de 0.1, acurácia como critério de escolha 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. Já para a associação as regras foram criadas considerando um suporte mínimo de 0.122 e uma confiança superior a 0.9. Esses números foram definidos com base em testes e análises dos resultados preliminares na perspectiva dos gestores da empresa.

Referências

- AUGUSTO JUNIOR, R. G.; Descoberta de conhecimento para identificação de fatores que influenciam o desempenho discente: diferença entre cursos de direito e engenharia civil. 2015. 158 f. Dissertação (Mestrado em Computação Aplicada) – Universidade do Vale do Itajaí (UNIVALI), Itajaí, 2015.
- MEHTA, N., STEINMAN, D., MURPHY, L. Customer success: how innovative companies are reducing churn and growing recurring revenue. 1 edition ed. New York: Wiley, 2016.
- SEBRAE, CNPQ. Comércio e Serviços. Cadernos de inovação em pequenos negócios, v. 2, p. 287, 2014.