

Aplicação Das Técnicas De Mineração De Dados E Clusterização Para Redução De Inadimplência Em Uma Empresa De Saneamento: Um Estudo De Caso

Daniel Felix Lamonica, Raquel Coutinho da Silva, Luiz Carlos Brasil de Brito Mello

Resumo: Este estudo tem como objetivo demonstrar a utilização de um software de mineração de dados capaz de mapear, analisar e gerar uma nova perspectiva para a questão da inadimplência em uma concessionária de saneamento básico, agrupando e concatenando os clientes específicos, viabilizando assim a proposta de um plano de ação personalizado com o que foi observado nas análises. Com um banco de dados dos clientes inadimplentes da empresa descrito no estudo de caso, foi possível clusterizar e analisar esses clientes, gerando um plano de ação visando a redução da inadimplência acumulada na empresa. O método para clusterização foi a utilização do software Power BI, especializado em trabalhar com grandes bancos de dados e capaz de gerar informações precisas para a tomada de decisão. O resultado da mineração de dados apresentou uma forte tendência dos clientes em se manterem inadimplentes, em paralelo foi possível observar como o modelo de cobrança da empresa é ineficiente e pouco ativo.

Palavras chave: mineração de dados, clusterização, inadimplência, plano de ação, banco de dados

Applying Data Mining and Clustering Techniques to Reduce Defaults in a Sanitation Company

Abstract: The study aims to use a data mining software capable of mapping, analyzing and generating a new perspective on the issue of non-payment in a basic sanitation concessionaire, grouping and concatenating the specific clients, thus making possible the proposal a personalized action plan with what was observed in the analyzes. With a client's database described and explained in the case study, it was possible to cluster and analyze these clients and generate a plan of action aimed at reducing default rate in the company. The method for clustering was Power BI software, specialized in large databases and capable of generating valuable information for decision making. The results of the data mining showed a strong tendency of the customers to remain in default, in parallel it was possible to observe how the collection model of the company is inefficient and not very active.

Key-words: data mining, clustering, non-payment, action plan, database

1. Introdução

A indústria de saneamento básico cumpre uma função primária e vital para toda e qualquer população. No Brasil, o acesso ao saneamento é assegurado para todos através da Lei nº. 11.445/2007, envolvendo a gama de serviços, infraestrutura e instalações operacionais que contemplam abastecimento de água, coleta de esgoto, limpeza urbana, drenagem, manejos de resíduos sólidos e de águas pluviais (Agência Nacional de Águas, 2017). Entretanto, é senso comum que há enorme precariedade nesse setor, impactando a maior parte dos brasileiros.

O gráfico 1 mostra a evolução quanto ao acesso da população brasileira à água tratada e captação de esgoto entre os anos de 2007 e 2015, fornecida pelo Instituto Trata Brasil:

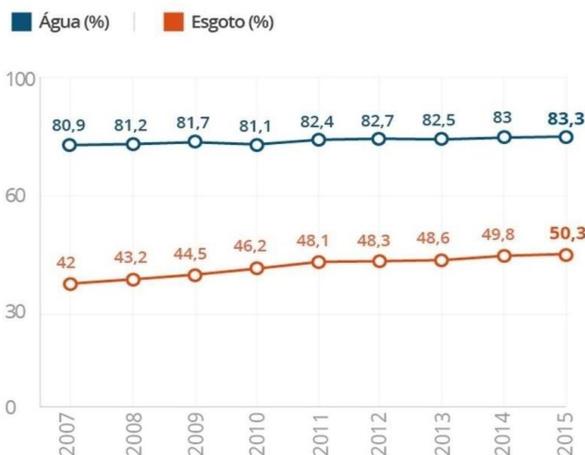


Gráfico 1: Acesso à população brasileira à água e esgoto
Fonte: Trata Brasil (2017)

O baixo desempenho em coleta de esgoto vai afetar diretamente a saúde da população impactada. De acordo com a Organização Mundial de Saúde (OMS) e também da Organização Pan-americana de Saúde (OPS), 65% dos que ocupam os leitos dos hospitais são pessoas acometidas por doenças provenientes de falta de saneamento. A mortalidade infantil, em grande parte dos casos, é gerada pela ausência de saneamento, em especial por diarreias infecciosas (TRIGUEIRO, 2005).

Segundo o SPC Brasil, a quantidade de inadimplentes no Brasil ultrapassa a marca das 61 milhões de pessoas. Dos débitos em aberto registrados nos cadastros da instituição, cerca de 8% são ligadas ao setor de Água e Luz. Pode-se estimar então que há cerca de 2 milhões de pessoas físicas em débito com o referido setor somente na região Sudeste (Confederação Nacional de Dirigentes Lojistas, SPC BRASIL, 2018).

O presente estudo está focado em analisar como clientes inadimplentes de uma concessionária de água e esgoto podem ser aglomerados (clusterizados) através de uma ferramenta de Data Mining. Como Côrtes *et al.* (2002) definem, *Data Mining* é uma cooperação entre o homem e a máquina, com o objetivo final de se extrair padrões e avaliar como variáveis se relacionam em grandes bancos de dados.

2. Setor de saneamento ambiental

De acordo com a Organização Mundial de Saúde (OMS) (Organização Mundial de Saúde, 2018), saneamento é o controle de todos os fatores do meio físico do homem, que são ou podem vir a ser nocivos para o bem-estar físico, mental e social. Ou seja, saneamento é considerado o conjunto de ações socioeconômicas que buscam alcançar Salubridade Ambiental. Esse termo pode ser entendido como o nível ambiental que possibilite a prevenção de doenças contraídas pelo meio ambiente e também melhorar as condições benéficas à saúde da população em geral (SÃO PAULO, 1999).

Além disso, a OMS apresenta Saneamento Ambiental em quatro momentos:

- Abastecimento de água: a disponibilidade de água de qualidade é uma condição indispensável para a própria vida e mais que qualquer outro fator, a qualidade da água condiciona a qualidade de vida. Por isso deve se levar em consideração o relacionamento íntimo entre água e saúde.

- b) Esgoto sanitário: os dejetos gerados pelas atividades humanas, comerciais, e industriais necessitam ser coletados, transportados, tratados e dispostos mediante processos técnicos, de forma que não gerem ameaça à saúde e ao meio ambiente.
- c) Resíduos Sólidos: esses resíduos gerados em determinadas regiões vêm se agravando como consequência do acelerado crescimento populacional, concentração das áreas urbanas, desenvolvimento industrial e mudanças de hábitos de consumo.
- d) Drenagem: a retenção da água na superfície do solo pode propiciar a proliferação dos mosquitos responsáveis por doenças como dengue e malária, além de outros vetores nocivos. Isso ressalta a importância da drenagem quanto ao nível de saúde populacional.

Em relação ao presente estudo, a concessionária onde será aplicado o estudo de caso presta os serviços somente relacionados ao abastecimento de água à população e a captação e tratamento do esgoto gerado.

3. Mineração de Dados (*Data Mining*)

Mineração de dados (em inglês, *data mining*) é a busca por informações que tenham algum valor para o seu usuário em grandes bancos de dados (WEIS, 1999). Thuraisingham (1999) define *data mining* como sendo o processo de aplicação de várias consultas para extrair essas informações úteis, através de padrões e tendências, que geralmente estão subliminares por conta da vastidão de dados.

Um conceito errado sobre mineração de dados é o que considera os sistemas de mineração como sendo capazes de minerar e gerar os dados de forma automática e os valores e conceitos mais úteis em uma extensa base de dados sem que haja a intermediação humana (HAN *et al.*, 2011). O ser humano cumpre papel fundamental, direcionando o sistema para agregar e associar os dados cruciais e que cirurgicamente vão gerar os resultados esperados.

A mineração de dados compõe diversas tarefas, sendo utilizadas da forma como se deseja identificar o padrão, e são ditas como descritivas ou preditivas (KANTARDZIC, 2011). Ambas as classificações serão melhor detalhadas nos tópicos a seguir.

a) Análise Descritiva

A Análise Descritiva, de acordo com Côrtes *et al.* (2002), significa a investigação nos dados que procura características relevantes, muitas vezes escondidas e não óbvias ao usuário, como analisar os dados, principalmente pelo seu aspecto de qualidade, para ratificar os resultados obtidos na mineração, ou seja, o conhecimento que foi gerado. A análise é subdividida em Análise Prévia e Descobrimto.

A análise prévia consiste em avaliar uma base de dados com o objetivo de encontrar possíveis anomalias e resultados raros que possam impactar os produtos da mineração de dados. Descobrimto consiste em analisar uma base de dados buscando padrões que podem estar obscuros, sem que exista uma hipótese levantada previamente (CÔRTEZ *et al.* 2002). Ambas as análises possuem sub-funcionalidades, que funcionam como técnicas adequadas para diferentes bases de dados e objetivos.

b) Análise Preditiva

A Análise Preditiva investiga os dados procurando inferir resultados a partir dos padrões resultantes na análise descritiva (descrita no tópico anterior), ou seja, prognosticar como irá se comportar um conjunto de dados novo. Berry e Linoff (1997), Thuraisingham (1999) e Han

et al (2011) subdividem essa análise da seguinte forma:

- Estimação: prediz algum valor, de acordo com um padrão pré-estabelecido. Por exemplo, em uma grande base de dados, sabendo as despesas e idade de uma pessoa nessa base, pode-se estimar seu salário (CÔRTEZ, 2002).
- Classificação: prediz um valor para uma variável categórica, ou seja, classifica um elemento do seu banco de dados em grupos pré-definidos. Como exemplo, uma instituição financeira que classifica clientes como oferecedor de risco ou não para se obter um empréstimo (CÔRTEZ, 2002).

4. Clustering ou Agrupamento

Segundo Han *et al* (2011), a análise de *cluster* é o processo de particionamento de um conjunto de dados em subconjuntos. Cada subconjunto é um *cluster*, de tal forma que os objetos em um *cluster* são semelhantes uns aos outros, mas diferem de objetos de outro subconjunto. Denomina-se como *clustering* o conjunto de agrupamentos resultados de uma análise. Neste contexto, diferentes métodos de *cluster* podem gerar diferentes agrupamentos no mesmo conjunto de dados. O particionamento mais refinado é feito através de algoritmos de *clustering*, e não diretamente por humanos (HAN *et al.*, 2011). Por levar a descobertas não antes propostas é que o método de *clustering* é considerado tão útil.

Tratando-se de mineração de dados, essa análise possibilita o uso de uma ferramenta autônoma para obter percepção da distribuição de dados, enxergar as características dos *k clusters* obtidos e focar em um conjunto particular de *clusters* para focar a análise (HAN *et al.*, 2011).

Existem muitos algoritmos capazes de realizar agrupamentos. Han *et al* (2011) afirmam que é difícil fornecer uma categorização de métodos de agrupamento. No entanto, é útil apresentar uma imagem relativamente organizada de métodos de agrupamento. Em geral, os principais métodos de agrupamento podem ser classificados nas seguintes categorias: Métodos de partição, métodos de hierarquia, métodos baseados em densidade e por fim métodos baseados em grades. A tabela abaixo resume as principais características de cada método.

Método	Características Gerais
Partição	- Encontra aglomerados mutuamente exclusivos de forma esférica
	- Baseado em distância
	- Pode usar média para representar o centro do <i>cluster</i>
	- Efetivo para conjuntos de dados de tamanho pequeno a médio
Hierárquico	- <i>Cluster</i> é uma decomposição hierárquica (ou seja, vários níveis)
	- Não corrige mesclagens ou divisões erradas
	- Pode incorporar outras técnicas como microcluster
Baseado em densidade	- Pode encontrar <i>clusters</i> de forma arbitrária
	- <i>Clusters</i> são regiões densas de objetos no espaço que são separados por regiões de baixa densidade
	- Densidade de <i>cluster</i> : cada ponto deve ter um número mínimo de pontos dentro de seu "bairro"
	- Pode filtrar <i>outliers</i>
Baseado em grade	- Usa uma estrutura de dados de grade de multirresolução
	- Tempo de processamento rápido (normalmente, independentemente do número de objetos de dados, ainda dependentes do tamanho da grade)

Quadro 1: Métodos de agrupamento

Fonte: Han *et al.* (2011)

Após a escolha do método mais condizente com a necessidade, precisa-se então avaliar os *clusters* gerados. A avaliação de *cluster* verifica a viabilidade da análise de *clustering* em um

conjunto de dados e a qualidade dos resultados gerados por um método de agrupamento. As tarefas incluem avaliar tendência de agrupamento, determinação do número de *clusters* e medição de *clustering* qualidade.

5. Estudo de Caso

a) Empresa estudada

A empresa onde foi aplicado o estudo de caso atua na cidade de Niterói – Rio de Janeiro desde 1998 como concessão privada prestadora de serviços de abastecimento de água, coleta e tratamento de esgotos. O grupo em que a concessionária pertence, está presente em mais de 13 cidades na região Sudeste, atende mais de 4 milhões de habitantes e por isso é líder no referido setor. Além disso, o grupo é 100% nacional e possui fortes parcerias com prefeituras e outras empresas privadas para alcançar metas de universalização do serviço de saneamento nos próximos anos.

A situação da inadimplência apresentada como problema focal deste estudo implica não só nos indicativos comerciais como nos resultados financeiros, sendo um assunto de interesse para ambas as áreas. Com o atual momento de melhoria do fluxo interno de informações entre as duas áreas, um projeto que atue em um impasse na interseção das áreas está de total acordo com a conduta da empresa.

b) Inadimplência na empresa

O setor de cobrança é responsável pelo controle ativo da inadimplência, utilizando mecanismos de contato direto com o cliente para a que a dívida seja sanada. Esses mecanismos são: *e-mails*, mensagens para o telefone celular e ligações. A empresa também possui um plataforma *on-line* em seu *site*, permitindo o parcelamento das faturas em aberto do cliente e tornando o saldo da dívida mais acessível. Devido a essa forte atuação do setor, a empresa recebe por mês uma média de 700 mil reais de débitos negociados em meses anteriores ao vigente.

Após o setor de cobranças fazer o levantamento dos dados de clientes inadimplentes, cabe ao setor de planejamento transformar esse banco de dados em números gerenciais. Com isso, são calculados indicadores vitais, como por exemplo: Inadimplência acumulada nos últimos 12 meses; Inadimplência acumulada desde o começo da concessão; Evasão da receita por inadimplência. Esses indicadores são expressos em relatórios mensais juntamente com outros indicadores financeiros/operacionais. Por fim, os dados já tratados são enviados a Superintendência Financeira, que juntamente com a Diretoria Administrativa-Financeira tomam as medidas cabíveis e fazem uma análise mais acurada, procurando identificar causas e possíveis soluções.

A diversificação e eficácia dos meios de cobrança é um ponto focal da área devido a expressiva taxa de dívida ativa. A inadimplência é tida como um importante indicador. É calculada dividindo-se o saldo de contas recebidas no mês pelo faturamento original (o que realmente foi consumido de água e esgoto tratado). Esse indicador ao final de 2017 atingiu um patamar de 5% no ano. Ou seja, o faturamento real do ano foi decrescido em cerca de R\$ 20 milhões de reais devido a esse problema.

c) Banco de Dados

Como o presente trabalho trata de uma mineração de dados em estado cru, surge a necessidade de se obter um banco de dados onde as informações de cada cliente inadimplente estejam de forma não gerencial. Ou seja, não tenham sido processadas ou melhoradas. O banco de dados deve conter a maior quantidade de informações possíveis para que o *software* possa fazer uma análise mais acurada. Algumas informações no banco de dados devem ser adaptadas para que o *software* possa fazer a clusterização.

A partir dessa necessidade, o banco de dados será obtido juntamente com o setor de cobranças, onde o mesmo possui uma base de dados cadastral utilizada para entrar em contato com os clientes inadimplentes e realizar a abordagem da negociação de dívidas. Esse banco de cadastro será conjugado com um banco de dados que possua as informações de consumo dos clientes. Isto é, o que foi consumido de água e tratamento de esgoto e não devidamente quitado. Com isso, pode-se obter uma modelagem conceitual entidade-relacionamento do banco de dados:

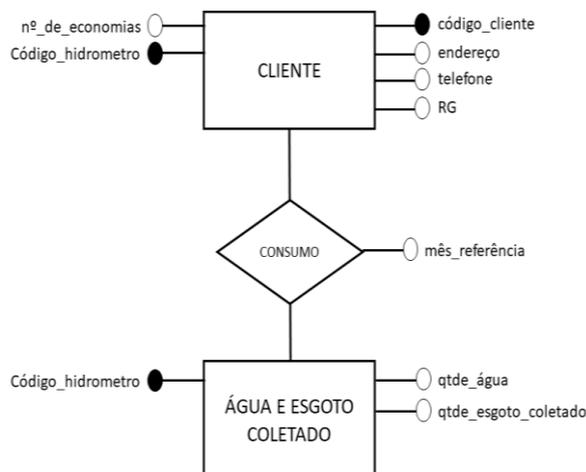


Figura 1: Modelo Conceitual Entidade-Relacionamento do Banco de Dados

Fonte: Autores (2018)

O modelo conceitual de relacionamento representado acima traduz como o banco de dados usado nesse estudo está estruturado. No banco existem duas entidades: o cliente e água e esgoto. Uma entidade pode ser entendida como um banco de dados possuindo instâncias (linhas) e atributos (colunas). Essas duas entidades conversam através do relacionamento consumo. Desta forma, está criada uma ligação entre as duas entidades com a quantidade de água e esgoto consumida por cada instância de cliente. Este relacionamento de consumo deve possuir um mês referência, ou seja, o que foi consumido em um mês específico.

A entidade de Cliente contém atributos específicos, que definem o cadastro de um cliente no sistema. Os atributos de código_cliente e código_hidrômetro identificam uma instância de cliente, ou seja, cada cliente só pode possuir um código de hidrômetro vinculado a ele e um hidrômetro está relacionado a um e somente um cliente. O atributo nº_de_economias diz respeito a quantidade de casas ligadas àquele hidrômetro (por exemplo, uma residência pode possuir casas anexadas, sobrados, etc. e ter somente um hidrômetro vinculado)

A entidade Água e Esgoto Coletado traduz a quantidade de água e esgoto consumida e é identificada por um código_hidrômetro. Isso significa que o padrão de consumo estará vinculado a um único hidrômetro, por consequência ligado a somente um cliente. Tratando esse consumo em um mês específico, concreta-se o relacionamento entre ambas as entidades e forma o escopo do banco de dados utilizado nesse estudo.

d) O software e a clusterização

O software utilizado para realizar a mineração de dados e as outras análises é o Power BI. O Power BI, desenvolvido pela Microsoft, é uma gama de softwares, aplicativos e conectores que trabalham em conjunto para relacionar dados e transformá-los em informações visualmente envolventes e interativas. Podem ser utilizados grandes bancos de dados ou até mesmo dados básicos em uma planilha Excel nessa ferramenta, permitindo que o usuário

conecte as fontes de dados, gere um conhecimento de forma lúdica e possa compartilhar isso com inúmeros outros usuários.

Após o Power BI puxar as informações de diversas fontes de dados (seja do sistema financeiro, do sistema de estoque etc.), é possível cruzar essas informações e publicar *dashboards* dinâmicos para diferentes usuários/funcionários. Com isso, diversas pessoas terão acesso rápido a informação através de diferentes plataformas.

Para a clusterização do banco de dados, o *software* utiliza um método de partição dos dados, por essa técnica ser baseada em distância. O número de *clusters* é definido pelo número de bairros (56). Assim sendo, os *clusters* são mutuamente exclusivos entre si (cada cliente pertence somente a um bairro) e a densidade do agrupamento é determinada pelo volume da dívida total de todos os clientes pertencentes ao grupo.

6. Resultados da clusterização e levantamentos da base

Para a clusterização, foi considerada a base de dados com todos os clientes inadimplentes da empresa desde o início da prestação de serviço na cidade (1999). Foi utilizado o método de partição para o agrupamento, tendo em vista que a formação dos *clusters* é a limitação do próprio bairro do município, caracterizando assim um critério de distância entre os clientes para que se formem os grupos. Niterói totaliza 56 bairros em seu território. A figura 2 apresenta o resultado do agrupamento, levando em consideração o volume absoluto de contas vencidas a receber.



Figura 2: Resultado da clusterização dos clientes inadimplentes por bairro de Niterói

Fonte: Autores (2018)

Os círculos vermelhos representam proporcionalmente a soma total dos valores das contas vencidas dentro de cada bairro (*cluster*). Sendo assim, pode-se concluir que alguns bairros se destacam pelo alto volume da dívida. Destacando os 5 principais bairros neste quesito, pode-se citar: Fonseca, Centro, Largo da Batalha, Barreto e Engenhoca. O bairro Centro deve ser tratado de forma singular, tendo em vista que é um agrupamento com altíssima concentração de clientes comerciais, que possuem débitos mais elevados e com isso gerar essa grande discrepância quando comparado com os outros bairros da cidade (todos possuindo uma relação residência/comércio acima de 80%). Além disso, grande parte desses clientes comerciais do Centro são tratados a parte por apresentarem processos judiciais em andamento, o que impede qualquer ação da empresa para recuperar o ativo da dívida até o processo judicial ser concluído.

Também deve-se levar em consideração o fato de que um cliente pode ter mais de uma conta em aberto.

Bairro	Cientes inadimplentes	Contas vencidas	Média de contas por cliente
FONSECA	5.639	312.042	55
MORRO DO ESTADO	1.326	235.516	178
LARGO DA BATALHA	2.893	228.719	79
BARRETO	2.881	204.320	71
ENGENHOCA	2.943	166.631	57
SAO FRANCISCO	1.633	118.557	73
CANTAGALO	1.249	97.461	78
CARAMUJO	1.246	90.932	73
JURUJUBA	999	80.602	81
ITITIOCA	905	71.473	79

Tabela 2: 10 bairros com maior número de contas vencidas

Fonte: Autores (2018)

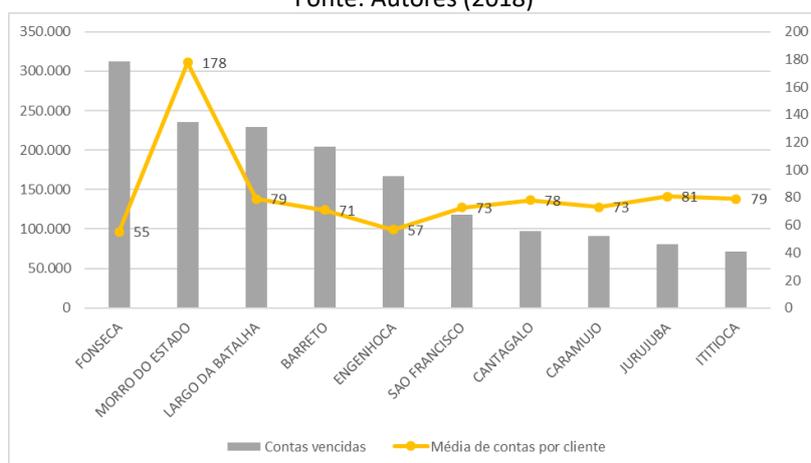


Gráfico 2: Número de contas vencidas por região e média de contas por cliente

Fonte: Autores (2018)

A Tabela 2 e o Gráfico 2 elucidam os 10 primeiros bairros no quesito número de contas vencidas. As quantidades de contas desses 10 bairros representam mais de 60% de toda a cidade de Niterói. A partir desses números, comparando-se com a quantidade absoluta de clientes inadimplentes por região, obtém-se a média de contas em aberto por cliente inadimplente. Esse índice se mostra bem elevado para todas as regiões, e a média de toda a cidade de Niterói chega ao patamar de 50 contas vencidas por cliente inadimplente.

Analisando mais profundamente as análises anteriores, pode-se inferir que um denominador comum entre todos os bairros que possuem maior número de contas vencidas é o fato da forte presença de comunidades carentes nessas regiões. O Morro do Estado, em especial, possui o maior índice de conta por cliente por ser uma das maiores comunidades carentes de Niterói. As comunidades representam uma variável desafiadora para mitigar a questão da inadimplência. O contato direto (físico ou virtual) com o cliente residente de comunidade carente é extremamente restrito, por questões básicas de segurança. A empresa adota uma conduta muito cautelosa, e por isso não há o corte de fornecimento em áreas consideradas de alto risco para os funcionários operacionais.

Outro levantamento de base relevante para o estudo é o aging das contas vencidas. O conceito interno de *aging* é o envelhecimento da conta vencida. Ou seja, há quanto tempo a data de vencimento daquela conta já expirou e ela continua sem quitação. Tratando-se das regiões com número de contas vencidas significativo (acima de 20 mil), se tem as seguintes médias de envelhecimento das contas.

Bairro	Média de Aging	
	em dias	em anos
FONSECA	3.034	8,43
MORRO DO ESTADO	2.910	8,08
LARGO DA BATALHA	2.771	7,70
BARRETO	3.210	8,92
ENGENHOCA	3.086	8,57
SAO FRANCISCO	2.852	7,92
CANTAGALO	2.689	7,47
CARAMUJO	3.439	9,55
JURUJUBA	2.457	6,83
ITITIOCA	2.591	7,20
BADU	2.527	7,02
BALDEADOR	3.265	9,07
ILHA DA CONCEICAO	2.565	7,13
VICOSO JARDIM	3.046	8,46
MACEIO	2.825	7,85
SAO LOURENCO	3.247	9,02
SANTA ROSA	2.853	7,93
CHARITAS	3.371	9,36
ICARAI	3.250	9,03
CUBANGO	3.016	8,38
PIRATININGA	2.816	7,82
CENTRO	3.255	9,04
SAPE	3.562	9,89
ITAIPU	2.888	8,02
INGA	3.224	8,96
PONTA DA AREIA	3.067	8,52
VITAL BRASIL	2.860	7,94
CACHOEIRA	2.852	7,92
VIRADOURO	3.029	8,41

Tabela 1: Média de envelhecimento de contas das principais regiões inadimplentes de Niterói
 Fonte: Autores (2018)

Analisando a Tabela 3, tem-se um valor mínimo para média de *aging* de 6,83 anos, e o máximo correspondente a 9,55 anos. A média total para as regiões acima é de 8,29 anos. O indicador mostra variância muito baixa, o que indica uma tendência muito forte dos clientes para se manterem inadimplentes ao longo de todo o território municipal, havendo pouco incentivo e cobrança para que isso seja revertido.

6. Plano de ação

a) What (O que será feito?)

O projeto está focado em reduzir a evasão de receita oriunda da inadimplência dos clientes da empresa. Essa expectativa será alcançada reformulando todo o modelo de cobrança de dívida utilizado atualmente, fazendo com que todos os inadimplentes desde o início da prestação de serviço da empresa na cidade de Niterói (1999) sejam levantados e tenham seus status da dívida atualizados e documentados internamente.

	Objetivo	Status Atual	Prazo
Meta Principal	Reduzir a porcentagem de receita descontada por conta da inadimplência para 4,25%	4,45%	1 ano
Meta Secundária	Aumentar os contatos diretos com clientes via ligação e mensagem de texto em 500%	120 contatos/dia	4 meses

Quadro 6: Metas a serem atingidas com o projeto
 Fonte: Autores (2018)

b) Why (Porque será feito?)

O recebimento de muitos desses ativos não sendo desconsiderados do orçamento anual devido à falta de atuação ativa e eficiente do modelo de cobrança vigente. Além disso, há a necessidade da diminuição dos custos relativamente altos com o atual modelo de cobrança. Atualmente, a cobrança é feita de forma manual em quase sua totalidade, com o envio de

SMS e ligações. Com a padronização e automatização do processo de cobrança, haverá uma redução com envios de mensagens de cobrança e também com o uso da URA (que por sua vez possui custos elevados de manutenção e atualização).

c) Who (Quem será o responsável?)

O projeto de implementação do sistema será responsabilidade da área Comercial da concessionária, e terá o apoio das áreas de Faturamento, Recursos Humanos e Tecnologia da Informação.

d) Where (Onde será feito?)

O objetivo principal do projeto é reduzir a porcentagem de inadimplência na empresa. Para atingir esse objetivo, todo o modelo de cobrança feito pela empresa será reestruturado com a implementação de um sistema especializado. Por esse motivo, o projeto ocorrerá dentro do setor Comercial e de Faturamento da empresa.

e) How (Como será feito?)

A ação consistirá na implementação de um sistema especializado na recuperação de ativos (no caso, gerenciamento e cobrança efetiva dos devedores), com o intuito de melhorar o modelo de cobrança e estreitar o relacionamento com o cliente inadimplente. O sistema tem como principais características administrar e controlar as informações do devedor e da respectiva dívida, integração dos históricos completos, realiza contatos, simulações, cálculos, parcelamentos e a disponibilização de meios de pagamentos.

Um ponto importante desse sistema é a possibilidade de interlocução com outros sistemas de gerenciamento. A empresa já possui uma área e um sistema para a gestão dos ativos da empresa, e a sinergia entre o sistema novo e o que faz a gestão de todos os ativos presentes na empresa é um fator que facilita e muito a implementação e incorporação entre as áreas do projeto. Vale ressaltar que a gestão de ativos é feita pela área de Controladoria. Entretanto, o novo sistema previsto nesse plano de ação ficará a cargo da gerência Comercial/de Faturamento.

Tratando-se da melhora de cobrança prevista pelo novo sistema, haverá um acionamento de forma massiva e automática através de diversos canais (SMS, e-mail, URA). Essa cobrança multicanal pode ser feita a partir de qualquer filtro feito no sistema, levando em consideração qualquer variável do devedor e/ou sua dívida, tornando possível focar a cobrança em um perfil pré-definido de cliente. O sistema permite que os textos disparados por SMS e *e-mail* sejam totalmente personalizados para cada cliente ou grupo de clientes.

Com o aumento da abordagem de cobrança, será necessário aumentar o capital humano relacionado a *call center*. A tendência é que a demanda de ligações relacionadas a renegociação de dívida aumente. Será necessário que haja um corpo maior de funcionários para atendê-la. O sistema proposto também é capaz de gerir a operação de *call center*, fazendo o monitoramento *on-line* das ligações e das filas de cobrança, criando assim um histórico de todas as ações.

Mesmo aumentando o recurso humano da área de *call center*, o sistema tem uma interface direta com o cliente. Ele permite uma repactuação de dívida conforme as possibilidades de pagamento do cliente. Após a renegociação, ele emite e envia boletos de forma automática (podendo ser com envio de link via SMS e e-mail, ou correio). O sistema está integrado diretamente com o sistema de integração da área financeira. Isso quer dizer que pagamentos recebidos pelo autoatendimento disponibilizado serão incorporados diretamente pela área de contas a receber.

Além dos ganhos com as ferramentas de cobrança, o sistema também disponibiliza a extração de *dashboards*, criando a funcionalidade de levantamentos de informações da base

transacional para gestão em tempo real. A eficácia do novo modelo de cobrança poderá ser verificada e controlada em tempo real, o que facilitará o acompanhamento de todos os indicadores e acarretará na rapidez da tomada de decisão de acordo com os resultados obtidos.

7. Conclusão

O estudo em questão teve como principal motivação levantar uma possível solução para a questão da inadimplência instaurada na empresa, através da clusterização dos clientes e levantamentos e análises que gerassem um novo e específico conhecimento sobre o problema tratado. A revisão da literatura realizada foi capaz de integrar os três pilares deste trabalho: a inadimplência no setor de saneamento básico; a clusterização como técnica de *data mining* e descoberta de conhecimento; e a utilização de técnicas de qualidade para solução do problema.

Sendo o primeiro pilar a inadimplência, foi esta a motivação e questão principal de solução do estudo. O segundo pilar, que foi a mineração de dado, trouxe o método de análise necessário para entender e extrair informações valiosas da base de clientes inadimplentes. O terceiro pilar foi a proposta que interliga o primeiro e segundo pilar, trazendo uma solução para o problema baseado na técnica de análise computacional realizada.

No decorrer do trabalho, ao tomar conhecimento e entender de forma holística a base de dados, foi possível enxergar as outras formas de análise (além da clusterização por região proposta inicialmente) e trazer maior credibilidade e aderência a realidade ao plano de ação. Entender todas as informações que o banco de dados oferece foi vital para selecionar quais tipos de levantamentos trariam um conhecimento útil.

Referências

Agência Nacional de Águas, 2017. Saneamento. Disponível em: <<http://www3.ana.gov.br/portal/ANA/aguas-no-brasil/>> Acesso em: 10 jul. 18

BERRY, Michael J. A.; LINOFF, Gordon. "Data Mining Techniques for Marketing, Sales, and Customer Support"; John Wiley & Sons, Inc., 1997.

BIO, S. R. Sistemas de informação: um enfoque gerencial. 2. ed. São Paulo: Atlas, 2008.

CATARINO, MARINO HILARIO. Integrando Banco De Dados Relacional E Orientado a Grafos Para Otimizar Consultas Com Alto Grau De Indireção. Dissertação de Mestrado. USP, 2017.

Confederação Nacional de Dirigentes Lojistas, SPC BRASIL, 2018. Indicadores Econômicos. Disponível em: <https://www.poder360.com.br/wp-content/uploads/2018/01/indicador-de-Inadimplencia-PF_-dez-2017.pdf> Acesso em: 09 jul. 18

Confederação Nacional de Dirigentes Lojistas, 2018. Após crescimento no auge da crise, inadimplência com contas de água e luz recua no país, apontam SPC Brasil e CNDL. Disponível em: <<http://site.cndl.org.br/apos-crescimento-no-auge-da-crise-inadimplencia-com-contas-de-agua-e-luz-recua-no-pais-apontam-spc-brasil-e-cndl/>>. Acesso em: 11 jul. 2018

CÔRTEZ, S.; PORCARO, R.; LIFSCITZ, S. Mineração de dados – Funcionalidades, Técnicas e Abordagens. PUC-RIO, maio, 2002

HAN, J.; KAMBER, M.; PEI, J. Data mining: concepts and techniques. 3. ed. San Francisco: Morgan Kaufmann, 2011.

HAND, D.; MANNILA, H.; SMYTH, P. Principles of data mining. Cambridge: MIT Press, 2001.

KANTARDZIC, M. Data mining: concepts, models, methods and algorithms. 2. ed. New Jersey: John Wiley & Sons, 2011.

KORTH, H. F.; SILBERSCHATZ, A. Sistema de banco de dados. 2.ed. São Paulo: Makron Books/McGraw-Hill, 1994

MINISTÉRIO DA SAÚDE. Fundação Nacional de Saúde. Impactos na saúde e no Sistema Único de Saúde decorrentes de agravos relacionados a um saneamento ambiental inadequado. Brasília, 2010. Disponível em: http://www.funasa.gov.br/site/wp-content/files_mf/estudosPesquisas_ImpactosSaude.pdf. Acesso em: 10 jul. 2018

Organização Mundial da Saúde, 2018. Saneamento Ambiental. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=579:saneamento-ambiental&Itemid=596> . Acesso em 10 jul. 2018

SÃO PAULO. Secretaria de Recursos Hídricos, Saneamento e Obras. ISA – Indicador de Salubridade Ambiental. Manual Básico. São Paulo, Brasil,1999.

SILVA, Albertina; FARIAS Daysi; SILVA, Vanessa. Análise Das Melhorias No Processo: Uma Abordagem Em Uma Empresa De Fibra De Vidro. Xi Congresso Nacional De Excelência Em Gestão, 2015.

STEINER, M.T.A. et al. Abordagem de um problema Médico por meio do Processo de KDD com ênfase à Análise Exploratória dos Dados. Revista Gestão & Produção, v.13, n. 2, p. 325-337, 2006. Disponível em: < <http://www.scielo.br/pdf/%0D/gp/v13n2/31177.pdf>>. Acesso em: 10 jul. 2018

THURASINGHAM, B.; “Data Mining”; CRC Press, 1999

TRATA BRASIL, 2017; Benefícios Econômicos E Sociais Da Expansão Do Saneamento No Brasil. Disponível em: <<http://www.tratabrasil.org.br/datafiles/estudos/beneficios-ecosocio/relatorio-completo.pdf>>. Acesso em: 12/07/2018

TRIGUEIRO, André. Mundo sustentável: Abrindo Espaço na Mídia para um planeta em transformação, Editora globo S. A, São Paulo, 2º Ed. 2005.

WEIS, S.M.; INDURKHYA, N. Predict Data Mining. Morgan Kaufmann Publishers, Inc., 1999.