

SARA NUNES DE JESUS ALVES

**ANÁLISE DE APLICAÇÃO DE TÉCNICAS DE DATA SCIENCE PARA
CRM EM AMBIENTE BANCÁRIO**

Monografia apresentada ao PECE – Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para conclusão do curso de MBA em Tecnologia de *Software*.

São Paulo
2016

SARA NUNES DE JESUS ALVES

**ANÁLISE DE APLICAÇÃO DE TÉCNICAS DE DATA SCIENCE PARA
CRM EM AMBIENTE BANCÁRIO**

Monografia apresentada ao PECE – Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para a conclusão do curso de MBA em Tecnologia de *Software*.

Área de Concentração: Tecnologia de *Software*

Orientador: Prof. Dr. Jorge Rady

São Paulo
2016

FICHA CATALOGRÁFICA

Alves, Sara

Análise de Aplicações de Técnicas de Data Science para CRM em
Ambiente Bancário / S. Alves -- São Paulo, 2016.
60 p.

Monografia (MBA em Tecnologia de Software) - Escola Politécnica da
Universidade de São Paulo. PECE – Programa de Educação Continuada em
Engenharia.

1.Ciência de Dados 2.CRM I.Universidade de São Paulo. Escola
Politécnica. PECE – Programa de Educação Continuada em Engenharia II.t.

DEDICATÓRIA

*Dedico este trabalho aos meus pais
por sempre acreditarem em mim até
quando eu mesma não acreditava.*

AGRADECIMENTOS

À Deus.

Aos meus pais Alice Nunes e Ivo Alves, à minha avó Anelita Nunes, à minha irmã Milena Alves e à toda a minha família pela compreensão nos momentos ausentes e pelo amor incondicional.

Aos meus amigos Marcelo Moraes, Milena Lima, Karina Martins e Janaina Lopes por me incentivarem e me motivarem ao longo da criação desse trabalho.

Ao meu orientador Jorge Rady, pela atenção, compreensão, paciência, direcionamento e suporte.

Aos meus colegas de trabalho André Moehleck, Nelson Bicho e Haroldo Zago, que contribuíram com informações importantes para o desenvolvimento desse trabalho.

À Escola Politécnica da Universidade de São Paulo – EPUSP que proporcionou um ambiente rico em material acadêmico de qualidade para suporte à pesquisa.

Ao PECE – Programa de Educação Continuada em Engenharia pelo conteúdo do curso e excelentes professores.

RESUMO

Este trabalho apresenta análises de técnicas de *Data Science* aplicadas à rotinas de CRM no ambiente bancário com o intuito de comprovar sua eficácia com relação ao aumento de assertividade de campanhas criadas. O conceito de *Data Science* é apresentado através do ponto de vista de diferentes autores à respeito do tema. As técnicas de *Data Mining*, *Machine Learning*, a incorporação da Estatística e importância da Engenharia de *Software* são apresentadas com o intuito de embasar a análise da aplicação do *Data Science* em pontos específicos de numa rotina de CRM.

Palavras-chave: *Data Science*, *Data Mining*, *Machine Learning*, Estatística, Ambiente Bancário e CRM.

ABSTRACT

This paper presents analysis of *Data Science* techniques applied to CRM routines in the banking environment in order to prove its effectiveness in relation to the increase of campaigns created assertiveness. The concept of *Data Science* is presented through the point of view of different authors. Techniques of *Data Mining*, *Machine Learning*, the incorporation of Statistics and importance of *software* engineering are presented in order to base the analysis of the implementation of the *Data Science* at specific points of a CRM routine.

Keywords: Data Science, Data Mining, Machine Learning, Statistic, banking environment and CRM.

LISTA DE ILUSTRAÇÕES

	Pág.
Figura 1 - Passos do Processo KDD	24
Figura 2 - Camadas da Engenharia de <i>Software</i>	27
Figura 3 - Processo de CRM em grande banco brasileiro	45
Figura 4 - Processo de CRM com sugestões de <i>Data Science</i>	51

LISTA DE ABREVIATURAS E SIGLAS

BACEN	Banco Central do Brasil
BC	Banco Central
BCB	Banco Central do Brasil
BNDES	Banco Nacional de Desenvolvimento Econômico e Social
CMN	Conselho Monetário Nacional
CODATA	Committee on Data for Science and Technology
CRM	Customer <i>Relationship Management</i>
CVM	Comissão de Valores Mobiliários
ECRM	Ebusiness Channel <i>management Relationships Management</i>
FMI	Fundo Monetário Internacional
IASC	International Association for Statistical Computing
ISI	International Statistical Institute
KDD	Knowledge Discovery in Databases
PIB	Produto Interno Bruto
SELIC	Sistema Especial de Liquidação e Custódia
SUMOC	Superintendência da Moeda e do Crédito
URV	Unidade Real de Valor

SUMÁRIO

	Pág.
1. INTRODUÇÃO	11
1.1 MOTIVAÇÕES.....	11
1.2 OBJETIVO.....	12
1.3 JUSTIFICATIVAS	12
1.4 ESTRUTURA DO TRABALHO.....	12
2. DATA SCIENCE	14
2.1 ESTATÍSTICA	19
2.2 MACHINE LEARNING	20
2.2.1 Aprendizagem Supervisionada (Supervised Learning).....	21
2.2.2 Aprendizagem Não Supervisionada (Unsupervised Learning)	21
2.2.3 Classificação (Classification)	22
2.2.4 Regressão (Regression).....	22
2.2.5 Árvores de Decisão (Decision Trees)	22
2.2.6 Aprendizagem Profunda (Deep Learning)	22
2.3 DATA MINING	23
2.4 ENGENHARIA DE SOFTWARE.....	27
2.5 CONSIDERAÇÕES DATA SCIENCE	29
3. CRM E O SISTEMA BANCÁRIO	30
3.1 PROCESSO DE IMPLANTAÇÃO DE CRM	31
3.2 SISTEMA BANCÁRIO BRASILEIRO	36
3.3 O USO DE CRM NO AMBIENTE BANCÁRIO	40
3.4 CONSIDERAÇÕES DE CRM E O SISTEMA BANCÁRIO.....	46
4. ANÁLISE DOS RESULTADOS	48
4.1 PROVA DE CONCEITO PARA CASES DE CAMPANHAS DE CRM	48
4.1.2 Modelo de gasto de cartão de crédito.....	49
4.2 A APLICAÇÃO DO DATA SCIENCE PARA CRM NO AMBIENTE BANCÁRIO	50
4.2.1 Análises estatísticas	51
4.2.2 Técnicas de Machine Learning	52
4.2.3 Técnicas de Data Mining	53
4.3 CONSIDERAÇÕES DAS ANÁLISES DE RESULTADO.....	53
5. CONSIDERAÇÕES FINAIS	55
REFERÊNCIAS.....	56

1. INTRODUÇÃO

O sistema bancário tem sofrido grandes mudanças nos últimos tempos devido à evolução tecnológica. Além da já conhecida concorrência entre bancos, atualmente as *Fintechs* que são *startups* especializadas em soluções financeiras, têm ganhado espaço no mercado por conta da agilidade e simplicidade de contratação de produtos, serviços, fácil contato através da internet e baixo custo oferecido. Essas empresas estão mudando o conceito de banco atual através de ações que recriam o modo como as pessoas pagam contas, administram o próprio dinheiro e usam o cartão de crédito.

Clientes com preferências a favor da tecnologia estão migrando para essas empresas e os bancos tradicionais perdem espaço de mercado ao mesmo tempo em que as *Fintechs* ganham relevância.

Com isso, surge um movimento generalizado de ações que visam compreender as tendências do mercado e as preferências dos clientes para oferecer o melhor produto e serviço antes do concorrente a fim de evitar a migração de clientes atuais e conquistar novos clientes.

O CRM (*Customer Relationship Management*) é a metodologia usada por todos os segmentos de mercado para manter o relacionamento com o cliente e componente importante para esse desafio. Porém com o grande volume de dados disponíveis, é necessário recorrer à tecnologias como o *Data Science* para tornar o trabalho viável.

1.1 Motivações

O estudo de aplicação das técnicas de *Data Science* foi motivado pela identificação de redução de eficácia das campanhas de CRM em um grande banco brasileiro. Essas campanhas visam contactar os clientes para oferecer-lhes melhores produtos e serviços levando em consideração suas necessidades, que são identificadas através da análise não automatizada de dados e comportamentos apresentados pelos clientes durante a interação com os canais de comunicação com o banco.

Levando em consideração o potencial de exploração dos dados armazenados no banco, viu-se uma boa oportunidade de aplicar técnicas de *Data Science* para melhorar a assertividade das campanhas oferecendo produtos e serviços para os clientes com base na análise refinada e menos manual do perfil de cada cliente.

1.2 Objetivo

O objetivo deste trabalho é apresentar a aplicação de técnicas de *Data Science* em campanhas de CRM para melhorar a assertividade da escolha do público.

1.3 Justificativas

Com o uso de técnica de *Data Science* é possível conhecer melhor o perfil dos clientes e com isso identificar o melhor produto e/ou serviço levando em consideração informações como o melhor momento para se ofertar, suas preferências de mídias (vídeo, banner estático, abordagem do gerente via chat, telefone, etc), situação financeira, objetivos, etc.

De posse de todas essas informações, fica mais fácil refinar o público adequado para determinada campanha e dessa forma aumenta-se a probabilidade desse público selecionado aceitar o que está sendo oferecido.

1.4 Estrutura do Trabalho

O Capítulo 1 INTRODUÇÃO apresenta as motivações, o objetivo, as justificativas e a estrutura do trabalho.

O Capítulo 2 *DATA SCIENCE* apresenta o embasamento teórico das técnicas que compoem o *Data Science*.

O Capítulo 3 CRM E O SISTEMA BANCÁRIO apresenta o conceito geral do que significa CRM e suas aplicações, descreve a cronologia histórica do sistema bancário brasileiro e apresenta a rotina de uma área de CRM de um grande banco brasileiro e seus pontos de melhoria.

O Capítulo 4 ANÁLISE DOS RESULTADOS descreve uma proposta de melhoria dos pontos apresentados no capítulo anterior usando técnicas de *Data Science* com base em uma prova de conceito realizada por uma consultoria americana.

O Capítulo 5 CONSIDERAÇÕES FINAIS descreve brevemente a motivação inicial e a conclusão do estudo realizado.

REFERÊNCIAS relaciona as fontes utilizadas como embasamento teórico ao longo do trabalho.

2. DATA SCIENCE

O termo *Data Science* já existe há muitos anos. Esse termo foi empregado pela primeira vez em 1966 por Peter Naur um informático dinamarquês. Ele usou o termo Datalogia (*Datalogy*) que é sinônimo de *Data Science* para substituir Ciência da Computação por entender ser mais adequado. A partir dessa declaração, o termo Datalogia foi adotado em países escandinavos, como por exemplo, Dinamarca e Suécia.

Em 1962, John W. Turkey publicou em *The Future of Data Analysis* (O futuro da análise de dados): “Por muito tempo eu pensei que era um estatístico, interessado em inferências a partir do particular para o geral. Mas como eu tenho visto a matemática estatística evoluir, tenho motivo para perguntar e duvidar... Comecei a sentir que meu interesse principal em análise de dados... Análise de dados e as partes estatísticas envolvidas na análise de dados... Assumir as características da ciência ao invés das de matemática... Análise de dados é intrinsecamente uma ciência empírica...”

Em 1974, Peter Naur publicou o livro *Concise Survey of Computer Methods* (Pesquisa concisa de métodos de computação). O livro é uma pesquisa de métodos contemporâneos de processamento de dados que são usados em uma ampla gama de aplicações. Nesse livro Naur define o *Data Science* como: "A ciência de lidar com dados, uma vez que eles tenham sido estabelecidos, enquanto a relação do dado para o que ele representa é delegado para outros campos e ciências."

De acordo com o site da Associação Internacional de Estatística Computacional (*International Association for Statistical Computing* (IASC)), em 1977 a IASC foi estabelecida como uma sessão do ISI – Instituto Internacional de Estatística (*International Statistical Institute*). “A missão do IASC é conectar metodologia da estatística tradicional, tecnologia de computação moderna e o conhecimento de especialistas de domínio, a fim de converter dados em informação e conhecimento.”

Jonathan Berry publicou uma reportagem na revista *Business Week* em setembro de 1994 com a seguinte frase: "As empresas estão recolhendo montanhas

de informação sobre você, trituram os dados para prever quão propenso a comprar um produto você é e usam esse conhecimento para criar uma mensagem de *marketing* precisamente calibrada para você comprar o produto..."

Em 1996, Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth publicaram o artigo *From Data Mining to Knowledge Discovery in Databases*, dizendo: "Historicamente a noção de encontrar padrões úteis em dados tem sido chamada de diferentes nomes, incluindo *Data Mining*, extração de conhecimento, descoberta de informação, colheita de informações, arqueologia de dados e padrão de processamento de dados... No nosso ponto de vista, KDD (*Knowledge Discovery in Databases*) se refere ao processo global de descoberta de conhecimento útil a partir de dados e *Data Mining* refere-se a uma determinada etapa neste processo. *Data Mining* é a aplicação de algoritmos específicos para extrair padrões de dados... Os passos adicionais no processo KDD, como a preparação, a seleção de dados, limpeza de dados, a incorporação do conhecimento prévio apropriado e interpretação adequada dos resultados da mineração, são essenciais para garantir o conhecimento útil derivado dos dados. A aplicação cega de métodos de *Data Mining* pode ser atividade perigosa, facilmente levando à descoberta de padrões sem sentido e inválidas"

De acordo com o site Spring Link, em 1997 a revista *Data Mining and Knowledge Discovery* é lançada, a inversão da ordem dos dois termos em seu título refletem a ascendência do "*Data Mining*" como a forma mais popular para designar "a extração de informação de grandes bancos de dados".

Em 2001, William S. Cleveland publicou o artigo *Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics*. É um plano "para ampliar as principais áreas de trabalho técnico do domínio das estatísticas. Porque o plano é ambicioso e implica mudança substancial, o campo alterado será chamado "*Data Science*".".

De acordo com David R. Lide e Gordon H. Wood, em abril de 2002 a revista *Data Science Journal* é lançada pelo *Committee on Data for Science and Technology* (CODATA), publicando artigos sobre "a gestão de dados e banco de dados em ciência e tecnologia. O escopo da revista inclui descrições de sistemas de dados, a sua publicação na Internet, aplicações e questões legais".

De acordo com o site *Journal of Data Science*, em 2003, a revista de *Data Science* é lançada na Universidade de Columbia: "Por '*Data Science*' nós queremos dizer tudo que o que fazer com dados: coletar, analisar, modelar... no entanto, a parte mais importante são suas aplicações de todos os tipos. Essa revista é dedicada a aplicações de métodos estatísticos em geral... A revista *Data Science* fornecerá uma plataforma para que todos que trabalham com dados apresentem seus pontos de vista e troquem ideias".

Em maio de 2005, Thomas H. Davenport, Don Cohen, e Al Jacobson publicaram *Competing on Analytics*, um relatório de pesquisa da *Babson College Working Knowledge Research Center*, descrevendo "a emergência de uma nova forma de competição baseada no uso extenso de análise, dados e baseada em fatos de tomada de decisão... Ao invés de competir em fatores tradicionais, as empresas começaram a usar análises estatísticas e quantitativas e modelagem preditiva como elementos principais de concorrência".

Em julho de 2008, Alma Swan e Sheridan Brown publicaram: *The Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs*. Nesse artigo eles definem o cientista de dados como "pessoas que trabalham onde a pesquisa é realizada, ou no caso do pessoal do centro de dados, em colaboração próxima com os criadores do dado final, podem estar envolvidos na investigação criativa e análises, permitindo que outras pessoas trabalhem com dados digitais e desenvolvedores em tecnologia de base de dados".

Em março de 2009, Kirk D. Borne e outros astrofísicos publicam o artigo *The Revolution in Astronomy Education: Data Science for the Masses*. Nesse artigo eles defendem uma educação diferenciada para a próxima geração, enfatizado a necessidade de se ensiná-los a extrair inteligência dos dados para o sucesso das ciências, comunidades, projetos, agências, empresas e economias.

Em junho de 2010, Mike Loukides escreveu o artigo *What is Data Science*. No artigo além de incluir definições de diversos autores para o termo *Data Science*, ele descreve o perfil dos cientistas de dados que segundo ele, combinam empreendedorismo paciente, a vontade de construir produtos de dados de forma incremental, a capacidade de explorar e a capacidade de interagir além de uma solução. De acordo com ele, cientistas de dados são pessoas interdisciplinares. Eles podem lidar com todos os aspectos de um problema, desde a coleta inicial de dados

até dados condicionados para tirar conclusões. Eles podem pensar fora da caixa para chegar a novas maneiras de ver o problema, ou para trabalhar com problemas muito amplamente definidos.

Em maio de 2011, David Smith descreve em seu artigo: *Data Science: what's in a name?*, o crescente uso dos termos "*Data Science*" e "cientista de dados". Ele afirma que os termos começaram a ser usados com mais frequência há pouco mais de um ano, mas realmente decolaram desde então: "Muitas empresas já estão contratando "cientistas de dados" e conferências inteiras são executadas sob o nome de "*Data Science*". Mas apesar da adoção generalizada, alguns resistiram à mudança dos termos mais tradicionais, como "estatístico" ou "quant" ou "analista de dados".".

Em setembro de 2012, Tom Davenport e D.J. Patil publicaram na *Harvard Business Review: Data Scientist: The Sexiest Job of the 21st Century*. A partir de então, inúmeras variações, complementos e termos relacionados ao *Data Science* surgiram e conseqüentemente diversos artigos, livros e outros materiais foram criados à respeito do assunto.

Atualmente é muito difícil falar sobre *Data Science*, sem falar sobre tecnologias, princípios, processos e conceitos, como por exemplo, *Data Mining*, *Machine Learning* e Estatística.

Segundo Foster Provost e Tom Fawcett (2013), *Data Science* é um conjunto de princípios fundamentais que orientam a extração de conhecimento de dados. *Data Science* envolve princípios, processos e técnicas para compreender fenômenos através da análise (automatizada) de dados. Por último, o objetivo do *Data Science* é melhorar a tomada de decisão. Isso geralmente é o interesse direto de negócios.

Extrair conhecimento de uma quantidade grande de dados exige a aplicação de técnicas e uso de tecnologias e metodologias complementares. Sendo assim, para entender *Data Science*, é necessário entender também alguns desses conceitos.

No passado, as empresas tinham departamentos inteiros de especialistas em estatística, matemática e análise de sistemas que eram responsáveis por coletar, organizar e compreender os dados da organização. Com o passar dos anos, a

volumetria desses dados foi ficando cada vez maior e esse trabalho se foi se tornando cada vez mais difícil e ineficiente para ser executado de forma manual.

A tecnologia evoluiu e com essa evolução novas formas de capturar e ler esses dados facilitaram o trabalho desses profissionais.

Novas opções de interpretação desses dados também surgiram, conforme os profissionais experimentaram diferentes formas de cruzamento dos dados. Na maior parte dos casos, inclusive atualmente, o foco é estudar o comportamento do cliente para entender qual produto ou serviço se adequa melhor ao perfil do cliente. Ou até mesmo para antever cancelamentos ou fraudes.

Os cientistas de dados precisam ter características específicas para exercer esse papel. Não basta somente conhecer técnicas e ferramentas, esse profissional precisa ter o hábito de questionar, buscar além do que está sendo apresentado, ser intuitivo, criativo e claro, conhecer do negócio que está sendo analisado.

De acordo com Thomas C. Redman (2013) existem bons cientistas de dados e excelentes cientistas de dados. Um bom cientista de dados pode ajudar a encontrar relacionamentos em grandes quantidades de dados - muitas vezes ideias (*insights*) importantes que você não teria chegado de qualquer outra forma. Um excelente cientista de dados, por outro lado, desenvolve novas ideias a respeito do mundo. Eles certamente utilizam os dados para desenvolver essas ideias.

É muito importante ressaltar o perfil dos profissionais que trabalham com *Data Science*, inclusive os que não executam de fato as atividades, mas que de alguma forma estão envolvidos, porque para trabalhar com ideias extraídas de dados, é necessário compreender o pensamento analítico que nos leva a encontrar valor nesse tipo de informação.

A aplicação do *Data Science* pode ser explorada pelos mais variados tipos de negócio.

Os cientistas de dados constroem modelos matemáticos com o objetivo de explicar o comportamento de um determinado grupo de dados.

É possível, por exemplo, identificar a probabilidade de uma pessoa gostar de uma música levando em consideração as características de outras músicas que essa pessoa goste.

Se uma pessoa gosta de cinco álbuns de bandas de *rock* e três álbuns de *jazz*, é possível pesquisar formas de se obter mais informações sobre essas músicas, como o ano de lançamento, subgênero e os compositores.

Supondo que entre os álbuns que essa pessoa gosta, a maior parte das músicas foi composta entre 1984 e 1992, é razoável assumir que sugerir músicas atuais de pagode para essa pessoa tem menos chance de ajudá-la a descobrir novas músicas do que sugerir estilos próximos ao período de tempo de composição das músicas que essa pessoa já gosta. Uma forma de identificar e sugerir novas músicas pode ser através de uma investigação da recorrência de estilos nos gostos de diferentes pessoas que também gostem de músicas compostas no mesmo período em que a pessoa analisada gosta.

A análise do *Data Science* está ligada à contextos específicos. O sucesso da análise depende diretamente da qualidade dos dados que estão disponíveis para que as premissas que os modelos assumem sejam no mínimo razoáveis para o problema que precisa ser resolvido. É extremamente arriscado tentar inferir os gostos musicais de uma pessoa sabendo somente seu nome completo, data de nascimento e esporte favorito por exemplo. O fato de se produzir mais informações, armazená-las e processá-las aumentam a relevância dos dados.

Atualmente com a variedade e volume de informações é possível obter escolhas muito mais assertivas através da combinação de informações. Porém, é humanamente impossível fazer esse tipo de análise manualmente para cada cliente disponível na base de dados, e por isso o *Data Science* é tão útil. Com ele, é possível inferir essas informações através de inteligência artificial e *Machine Learning*, que são formas de “ensinar” o computador a inferir informações.

2.1 Estatística

Estatística é uma disciplina exata que através de teorias de probabilidade consegue explicar a frequência que um evento ocorre. Essa análise de frequência de eventos é usada para estimar a previsão de fenômenos futuros.

Segundo Boyer (1996), a palavra estatística surgiu da expressão em Latim *statisticum collegium* (palestra sobre os assuntos do Estado), originando o termo em língua italiana *statista*, que significa "homem de estado", ou político. Em alemão, *Statistiken* designa a análise de dados sobre o Estado. O sentido adotado atualmente surgiu no início do século 19.

Para Johnson e Leone (1964), a estatística pode ser dividida em três partes: 1) Estatística descritiva: Compreende a organização, o resumo e, em geral, a simplificação de informações, que podem ser muito complexas; 2) Probabilidade: Útil para analisar situações que envolvam acaso, como distribuições diamétricas, volumétricas, falhas em mecanismos, jogos de dados e cartas; e 3) Inferência: Diz respeito à análise e interpretação de dados amostrais.

Ainda segundo Johnson e Leone, na estatística os dados são coletados, analisados e interpretados. Existem métodos de obtenção, organização, resumo, interpretação e apresentação dos dados, além de ser possível também tirar conclusões com base nas características das fontes dos dados analisados para melhor compreensão das situações.

Para o campo computacional, com a evolução da tecnologia e o aumento na capacidade de processamento e armazenamento de dados, a Estatística foi se tornando cada vez mais relevante. Os modelos não-lineares, como redes neurais e árvores de decisão junto com algoritmos numéricos apropriados causaram um aumento do interesse na Estatística como forma de auxiliar na análise, processamento e interpretação da grande quantidade de dados gerada.

2.2 Machine Learning

Segundo Maria Carolina Monard e José Augusto Baranauskas (2003), *Machine Learning* é uma área de Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática.

Arthur L. Samuel (1959), define *Machine Learning* como “Um campo de estudo que dá aos computadores a habilidade de aprender sem serem programados explicitamente”.

Com *Machine Learning* são desenvolvidos algoritmos e técnicas que tentam reproduzir o funcionamento do cérebro humano e permitem ao computador aprender e tomar decisões baseadas em dados ao invés de apenas executar uma determinada tarefa. Alguns exemplos bem conhecidos de *Machine Learning* são as sugestões de filmes do Netflix, a playlist da semana apresentada pelo Spotify e a atual *timeline* do Facebook que prioriza a apresentação das notícias do interesse de cada usuário.

De acordo com Ian H. Witten e Eibe Frank (2005), existem alguns conceitos importantes do *Machine Learning* que precisam ser descritos:

2.2.1 Aprendizagem Supervisionada (*Supervised Learning*)

Ocorre quando o computador recebe um conjunto de dados pré-definido e com base nesses dados é treinado para tomar decisões precisas. Exemplo: Pode-se usar um conjunto de dados de avaliação de filmes para treinamento da *Machine Learning*, que tenha filmes marcados como bons, ruins ou neutros e assim treinar um classificador de análise de filmes.

2.2.2 Aprendizagem Não Supervisionada (*Unsupervised Learning*)

Ocorre quando o computador encontra automaticamente padrões e relações em um conjunto de dados. Exemplo: Análise de um conjunto de dados de tweets e agrupamento automático de tweets relacionados ao tema, sem que o *Machine Learning* possua qualquer conhecimento prévio sobre os dados.

2.2.3 Classificação (*Classification*)

Ocorre quando, com base numa entrada, o dado é rotulado. A Classificação é uma sub-categoria de Aprendizagem Supervisionada. Exemplo: Análise de uma imagem de um animal e classificação como gato ou cachorro.

2.2.4 Regressão (*Regression*)

A Regressão também é uma subcategoria de Aprendizagem Supervisionada e é usada quando o resultado que está sendo previsto é diferente de um simples “sim ou não”. Com a Regressão, é possível por exemplo responder às perguntas: “Quanto custa?” ou “Quantos existem?”.

2.2.5 Árvores de Decisão (*Decision Trees*)

Árvore de Decisão é uma ferramenta de apoio à decisão que usa um gráfico de árvore ou modelo de decisões e suas possíveis consequências. É possível usar Árvores de Decisão para representar visualmente um algoritmo.

2.2.6 Aprendizagem Profunda (*Deep Learning*)

A Aprendizagem Profunda é uma categoria de algoritmos de *Machine Learning* que usa redes neurais artificiais para gerar modelos. As técnicas de Aprendizagem Profunda são muito bem sucedidas na resolução de problemas de reconhecimento de imagem, pois possuem a capacidade de separar a imagem analisada em camadas e compreender os detalhes de cada camada individualmente. Exemplo: Na análise de uma foto, a primeira camada da Aprendizagem Profunda vê a foto como um todo identificando aspectos globais da foto. A segunda camada analisa regiões menores da foto e finalmente a terceira camada é responsável por analisar aspectos finos da foto como olhos, mãos, etc. É

assim que o Google Image consegue classificar as fotos que são carregadas no sistema.

2.3 Data Mining

Segundo Foster Provost and Tom Fawcett (2013), *Data Mining* é a extração do conhecimento do dado, via tecnologias que incorporam esses princípios.

Data Mining consiste em explorar muitos dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para identificar relacionamentos sistemáticos entre variáveis e consequentemente criar novos subconjuntos de dados. Com o *Data Mining* é possível identificar fatores e tendências-chave, descobrir regras, padrões e relacionamentos ocultos em grandes bancos de dados para auxiliar a tomada de decisões sobre estratégia e vantagens competitivas.

Data Mining é o meio pelo qual o *Data Science* encontra a resposta para as perguntas feitas pelo negócio. É através de algoritmos estruturados, matrizes, vetores, *softwares* para processamento de dados como Hadoop, que é uma plataforma de software para processamento de grandes massas de dados e MongoDB, que é um banco de dados orientado a documentos, que toda essa “mágica” acontece.

É possível usá-lo, por exemplo, para analisar comportamento de clientes a fim de descobrir formas de melhorar a retenção e aumentar a oferta de produtos. No mercado financeiro pode ser usado para pontuação de crédito e negociação e detecção de operações fraudulentas.

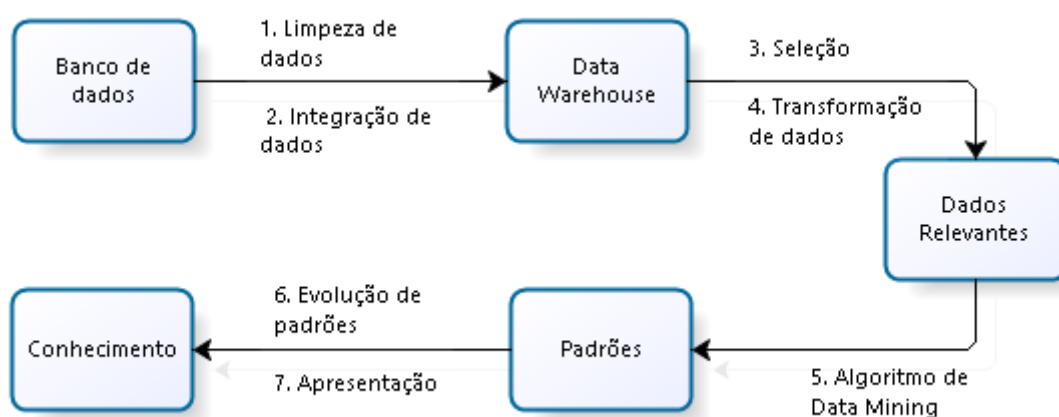
O *Data Mining* é composto por ferramentas e técnicas que através de algoritmos de aprendizagem e/ou estatística, é possível explorar um conjunto de dados, extraindo ou evidenciando padrões. Os padrões são importantes para auxiliar na descoberta de conhecimento.

O objetivo das técnicas do *Data Mining* é instrumentalizar a busca por padrões em um enorme volume de dados e executar essa busca em um prazo razoável. O que não seria possível se essa busca fosse feita manualmente.

Para isso, o *Data Mining* executa uma sequência de passos fundamentais a partir de fontes de dados que podem ser bancos de dados, transações, relatórios, logs de acesso, etc.

Esses passos consistem de uma limpeza e integração dos dados, que são dispostos em repositórios organizados (*Data Warehouses*) e a partir do *Data Warehouse*, é possível “refinar” o dado até que os padrões sejam descobertos.

Figura 1 - Passos do Processo KDD



Fonte: Sara Nunes, adaptado de Shivam Agarwal (2013)

De acordo com SHIVAM AGARWAL (2013), o processo de extração de conhecimento executado através do *Data Mining* é conhecido como KDD - *Knowledge Discovery in Database*, traduzido livremente como Extração de conhecimento de base de dados. Esse processo é composto por sete etapas: 1. Limpeza de dados, 2. Integração de dados, 3. Seleção, 4. Transformação de dados, 5. Aplicação de algoritmos de *Data Mining*, 6. Evolução de padrões e 7. Apresentação.

As quatro primeiras etapas são utilizadas para o pré-processamento de dados, esses dados são colocados em um formato para serem usados nas três etapas seguintes que são responsáveis por encontrar a informação escondida.

Limpeza de dados é a primeira etapa e serve para remover todo o ruído e outros dados inconsistentes da base de dados.

A segunda etapa é a Integração de dados, que serve para integrar os dados vindos de diferentes fontes. Esses dados limpos e integrados são armazenados em um *Data warehouse*.

Seleção de dados é a terceira etapa onde os dados mais relevantes para o *Data Mining* são selecionados.

A Transformação de dados é a quarta etapa, onde os dados relevantes são colocados num formato adequado.

A aplicação de algoritmo de *Data Mining* é a quinta etapa, onde as metodologias inteligentes são aplicadas nos dados para gerar conhecimento ou padrões.

Evolução de padrões é a penúltima etapa onde os padrões gerados na etapa anterior são refinados para que a informação relevante para tomada de decisão seja finalmente apresentada em um formato amigável na última etapa chamada de Apresentação.

Todo esse processo é interativo, sendo assim, é possível retornar à etapa anterior se for necessário. Essa reexecução da etapa anterior pode trazer novas hipóteses e descobertas e o executor do processo decide se continua, se inclui ou troca os atributos para avaliar novas hipóteses por exemplo.

Com o *Data Mining*, é possível encontrar informações relacionadas de diversas formas como por associações, sequências, classificação, aglomeração e prognósticos.

As informações relacionadas via associações estão ligadas à um único evento. Como por exemplo, um estudo de dados realizado em um supermercado pode revelar que na compra de sabão em pó, compra-se também amaciante de roupas em 65% das vezes, mas quando há uma promoção, o amaciante é comprado em 85% das vezes. Com essa informação, os gestores podem optar por decisões mais acertadas pois aprendem sobre a rentabilidade de uma promoção.

No caso das informações relacionadas via sequência, os eventos estão ligados ao longo do tempo. Dessa forma, é possível descobrir, por exemplo, que pessoas que compram uma bicicleta, em 65% das vezes compram um novo tênis no

período de duas semanas após a compra da bicicleta. Em 45% das vezes, um capacete também é comprado um mês após a compra da bicicleta.

Para classificação, as informações são reconhecidas em modelos que descrevem o grupo ao qual o item pertence por meio do exame dos itens já classificados e pela inferência de um conjunto de regras. Por exemplo, em companhias telefônicas existe uma preocupação muito grande com a perda regular de clientes. A classificação ajuda a identificar características de clientes que provavelmente cancelarão o serviço e oferecerá um modelo para ajudar a companhia a prever quem são esses clientes, de modo que seja possível elaborar antecipadamente campanhas especiais para reter esses clientes.

A aglomeração tem características semelhantes a aplicação de classificação quando os grupos ainda não foram definidos. Uma ferramenta de *Data Mining* é capaz de descobrir diferentes agrupamentos dentro da massa de dados. Por exemplo, um banco pode encontrar grupos de afinidades para cartões de crédito ou ao dividir a massa de dados em categorias de clientes com base na região do país em que moram.

Todas as aplicações mencionadas envolvem algum tipo de previsão, já os prognósticos as utilizam de um jeito diferente. Com base em uma série de valores existentes, os prognósticos preveem quais serão os valores seguintes. Por exemplo, com prognóstico é possível descobrir padrões que ajudam a estimar o valor futuro de variáveis com números de vendas.

As técnicas do *Data Mining* são muito poderosas e quando aplicadas apropriadamente são capazes de apresentar resultados extremamente relevantes. Porém, é importante que a equipe envolvida no processo conheça o contexto em que a informação está inserida e que o objetivo da análise também esteja claro para todos para que seja possível realizar a análise da forma mais adequada possível com a aplicação da técnica correta para cada situação.

2.4 Engenharia de Software

Os programas que rodam nos computadores diariamente são muito complexos. Cada parte do *software* é desenvolvida por equipes inteiras de especialistas nas tecnologias usadas em cada um deles e, não obstante, uma equipe pode ser formada por gente que não se conhece pessoalmente e que colabora de países diferentes. Além disso, com o crescimento da adoção das tecnologias, faz-se necessário pensar sobre como os serviços desenvolvidos devem continuar eficientes apesar do aumento da demanda. Isso é fundamental para que o produto final tenha alguma sobrevida. Por esse motivo, escrever programas de computador requer muito mais do que conhecer de programação. Segundo Roger S. Pressman (2011), a Engenharia de *Software* é o conjunto de práticas e padrões que facilita o desenvolvimento e a gestão dos produtos de *software* usados.

Figura 2 - Camadas da Engenharia de Software



Fonte: Sara Nunes, adaptado de Roger S. Pressman (2011)

Ainda de acordo com Roger S. Pressman (2011), a Engenharia de *Software* é uma tecnologia em camadas. Referindo-se à Figura 2, qualquer abordagem de engenharia (inclusive Engenharia de *Software*) deve estar fundamentada em um comprometimento organizacional com a qualidade. A gestão da qualidade total Seis Sigma e filosofias similares promovem uma cultura de aperfeiçoamento contínuo de processos, e é esta cultura que, no final das contas, leva ao desenvolvimento de abordagens cada vez mais efetivas na Engenharia de *Software*. A pedra fundamental que sustenta a Engenharia de *Software* é o foco na qualidade.

As três camadas que compoem a Engenharia de *Software* segundo Pressman, são: Processo, Métodos e Ferramentas.

A base para a Engenharia de *Software* é a camada de processos. Processo de Engenharia de *Software* é o que possibilita desenvolver *softwares* de forma racional e dentro do prazo. É uma sequência coerente de práticas usadas no desenvolvimento ou evolução de sistemas de *software*.

Essas práticas são compostas de atividades, ações e tarefas e são caracterizadas pela interação entre ferramentas, pessoas e métodos. Um processo não precisa ser necessariamente executado de maneira rígida. Ao contrário, um processo pode ser adaptado pelas pessoas envolvidas para que seja possível selecionar e escolher o conjunto apropriado de ações e tarefas.

Com Processo é possível definir uma metodologia para a entrega efetiva da tecnologia de *software*.

A camada de Métodos é responsável por direcionar como construir o *software*. Os Métodos são compostos por diversas tarefas, que incluem: comunicação, análise de requisitos, modelagem de projeto, construção de programa, testes e suporte.

As Ferramentas da Engenharia de *Software* proporcionam apoio automatizado ou semiautomatizado aos Métodos. Essa automatização não auxilia apenas no resultado final, é possível ter o apoio das ferramentas de automatização durante todo o processo de Engenharia de *Software*, passando pela análise de requisitos, modelagem, programação e testes.

Levando em consideração que computadores são ferramentas muito eficientes para analisar grandes volumes de dados, escrever programas de computador faz parte desse novo contexto tecnológico. Não ter uma forma automatizada de implementar os modelos matemáticos do *Data Science* que são usados para entender melhor as pessoas e melhorar a vida delas é como ter um médico, mas não dispor de instrumentos como medidores de pressão, monitores cardíacos, agulhas, remédios e outras ferramentas. E é por isso que a Engenharia de *Software* é tão importante para o *Data Science*.

2.5 Considerações Data Science

O *Data Science* já existe a mais de cinquenta anos e ao longo desses anos ele está em constante evolução e adaptação. Essa evolução acontece por conta da evolução dos demais setores da tecnologia e do mundo como um todo.

O emprego da Estatística, *Data Mining* e outros conceitos fundamentais para o *Data Science*, motivou a criação um novo perfil de profissional conhecido como cientista de dados. Esse profissional precisa ter conhecimentos que envolvem disciplinas distintas, promovendo uma nova categoria de profissionais que não saem “prontos” da universidade.

Um dos pilares do *Data Science* é a Estatística, sem o conhecimento de Estatística é impossível criar modelos que são os responsáveis pelas ideias de hipóteses e interpretações da leitura de uma base de dados.

Já com o *Machine Learning*, é possível distribuir o esforço de interpretar as informações geradas com o computador. Através do *Machine Learning* e dos resultados obtidos por ele, é possível entender conceitos que não seriam possíveis de forma manual.

O uso do *Data Mining* se faz necessário pelo tamanho da massa de dados que são trabalhadas atualmente. Suas técnicas são importantes para estruturar e classificar os dados a fim de facilitar o trabalho de aplicação das técnicas de extração de conhecimento.

Por fim, todo esse trabalho precisa estar numa estrutura sólida e segura. Esse é o papel da Engenharia de *Software*.

3. CRM E O SISTEMA BANCÁRIO

CRM é a sigla para *Customer Relationship Management*, que traduzido para o português significa Gestão de Relacionamento com o Cliente.

Para Plata, (2003) a Gestão de Relacionamento com o Cliente é uma estratégia cujo propósito é transformar os processos de negócios para conservar e conseguir mais clientes.

Segundo Swift (2001) o conceito de CRM não tem a ver com uma estratégia, ele define CRM como uma ferramenta administrativa para entender e influenciar o comportamento dos clientes, por meio de comunicações significativas para melhorar, além da retenção, as compras, lealdade e a lucratividade.

Thompson (apud GREENBERG, 2001) define CRM como um método para selecionar e administrar os clientes, buscando otimizar o valor a longo prazo. E ainda complementa dizendo que CRM requer uma filosofia e cultura empresariais centradas no cliente para dar suporte aos processos eficazes de *marketing*, vendas e serviços.

Segundo Pedron e Saccol (2009), o CRM é um conceito complexo que inclui três abordagens:

- Abordagem Filosófica: Guia o relacionamento da empresa com qualquer stakeholder.
- Visão Estratégica: Foca na ação e planejamento estratégico da empresa direcionado aos seus clientes, podendo gerar vantagem competitiva.
- Tecnologia da Informação: Provê condições estruturais para a execução do CRM, seu planejamento e suas interações e valores.

Levando em consideração as definições mencionadas, pode-se afirmar que CRM é uma metodologia de administração estratégica que com o auxílio de ferramentas automatizam funções de contato com o cliente para manter um bom relacionamento com o objetivo de aumentar o lucro da empresa.

No CRM, os dados dos clientes são coletados e armazenados. Com esses dados é possível entender qual a necessidade do cliente e antecipar suas necessidades. O CRM é voltado para o foco, aquisição, transação, atendimento, retenção e construção de relacionamento de longo prazo com os clientes.

Segundo Las Casas (2012, apud NICKELS; WOOD, 1999), no *Marketing Transacional* os vendedores focam trocas individuais e isoladas que promovem satisfações pontuais, tendendo a concentrar esforços na atração de novos clientes para trocas futuras.

Cobra (2009) define o *Marketing Relacional* como a criação, conquista e manutenção de clientes estabelecendo relacionamentos de longo prazo com os mesmos.

A origem do CRM está nas técnicas tradicionais de *Marketing* (*Marketing Transacional*) cujo foco é na venda e evoluiu para o *Marketing Relacional* cujo foco é no conhecimento, entrega constante de valor e fidelização dos clientes.

O CRM evolui as definições de *Marketing Relacional*, ao incorporar elementos de Tecnologia da Informação, tais como *DataWarehouse*, *Data Mining*, Automação de Força de Vendas, Canais de Relacionamento e suas variantes, para unificar esforços de relacionamento com o uso intensivo da tecnologia aplicada, segundo Cobra e Brezzo (2010).

3.1 Processo de implantação de CRM

Kotler (2005) afirma que antes da empresa iniciar um processo de implantação de CRM deve primeiramente, focar toda sua estrutura organizacional para o cliente. Pois, somente dessa forma, é possível utilizar o CRM corretamente. De nada adianta ter informação sobre os hábitos do cliente e não usar esses dados para atingir suas necessidades. Isso geralmente ocorre quando o foco da empresa está no produto.

Por isso o CRM coloca o cliente no centro dos processos de negócio. Ele é desenhado para perceber e antecipar as necessidades dos clientes atuais e

potenciais com o objetivo de supri-las da melhor forma. Essa é uma estratégia de negócio, que se apóia em soluções tecnológicas. Os *softwares* que auxiliam e apoiam esta gestão são normalmente denominados sistemas de CRM.

Com a implantação do CRM, a empresa tem a possibilidade de conhecer profundamente os seus clientes, o que precisam e em que perfil de consumidor se enquadra. Consegue criar respostas personalizadas, antecipando as suas vontades e respondendo de forma precisa aos seus desejos atuais. Com isso, é possível alcançar maior qualidade no contato com o cliente, por estar baseando-se em uma melhor e mais atualizada informação.

A mudança de cultura da empresa é também de extrema significância para o sucesso do processo, pois a torna mais proativa. Cada contato com o cliente fica mais significativo, assertivo e conseqüentemente decisivo para os negócios.

O papel da tecnologia é auxiliar na captura, organização de dados do cliente e fontes externas, consolidando esses dados de forma centralizada, geralmente criando um data warehouse.

A tecnologia também é responsável por integrar o *marketing* e as tecnologias de informação já existentes, de forma a prover para a empresa, meios eficazes e integrados de atender, reconhecendo e cuidando do cliente em tempo real. As aplicações de CRM transformam os dados centralizados em informação que permite a identificação do cliente e a compreensão do seu perfil.

No processo de implantação do CRM, é importante ter em mente o foco no cliente mencionado por Kotler (2005), bem como definir a abordagem que será feita ao cliente e definir um plano de comunicação interna. É necessário fazer um levantamento dos processos e documentações existentes, analisar e se necessário redesenhar os processos e criar documentações adicionais de apoio.

Com base nesse levantamento, a solução de informação e a estratégia de relacionamento com o cliente são escolhidas, bem como as ferramentas de apoio, por exemplo, *telemarketing*, canais virtuais, terminais de ponto de venda, etc.

Toda a empresa precisa estar envolvida e comprometida com o projeto, principalmente a gestão principal. O foco deve estar na mudança de processos.

A tecnologia deve ser escolhida de acordo com o tipo de negócio, a fim de atender as necessidades específicas do negócio. Deve-se analisar as funcionalidades realizadas e a arquitetura do sistema para se ter fundamento e base para apoiar os usuários finais.

A equipe de projetos precisa ter competências multidisciplinares para criar e garantir o estabelecimento de etapas do projeto e prioridades, a fim de integrar os sistemas de forma consistente.

Com o auxílio das ferramentas de sistema, implantadas como apoio ao CRM, é possível observar os seguintes pontos:

- Planejamento: Identificar todos os pontos críticos de relacionamento e a estrutura do fluxo de trabalho que resulta em um levantamento de pontos para automação
- *Marketing* relacional: Construir perfis de consumidor com base na identificação, segmentação, interação com os clientes.
- Produtos e serviços: Ênfase na gestão do conhecimento que resulta em redução de erros, aumento da qualidade e eficiência.
- Portal de *e-commerce*: Aumento da eficiência e atratividade para promover a divulgação dos produtos e serviços. Oportunidade de sondagens consumo e estudos estatísticos de mercado.
- Relacionamento: Aumento de valor ao relacionamento, com informações corretas e completas do que se deseja comprar ou contratar.
- Integração: Facilitar a integração de outros canais de comunicação, como telefone, fax, e-mail, etc. Essa integração facilita o enriquecimento da base de dados principal.
- Oportunidades de negócio: Com base na análise dos dados, é possível identificar padrões de comportamento econômico e de relacionamento, com isso é possível inferir eventuais formas de criar novos negócios.

Os pontos levantados dão suporte para a empresa angariar novos clientes e fidelizar clientes atuais, buscando atingir melhores níveis de satisfação através do melhor entendimento das suas necessidades e expectativas.

A junção de processos e sistemas de Gestão de Relacionamento com o Cliente possibilita o controle e conhecimento de informações sobre os clientes de forma estruturada e integrada através do acompanhamento e registro das interações realizadas com os clientes. Dessa forma, é possível consultar e comunicar diversas áreas da empresa que precisam desse tipo de informação para auxiliar no processo de tomada de decisão. Para isso, uma atividade importante da gestão de relacionamento com o cliente é o registro dos contatos de forma centralizada. Qualquer informação relevante para as tomadas de decisões devem ser catalogadas e analisadas periodicamente. Com isso é possível criar relatórios de gestão que servem de indicadores gerais da carteira de clientes.

Segundo Kotler (2005), existem basicamente quatro tipos de aplicações de CRM's:

- CRM Operacional: É a aplicação da tecnologia de informação para melhorar a eficiência do relacionamento entre os clientes e a empresa. Com o CRM Operacional é possível prever a integração de todos os produtos de tecnologia para proporcionar o melhor atendimento ao cliente.
- CRM Analítico: componente do CRM que permite identificar e acompanhar diferentes tipos de clientes dentro da carteira de uma empresa e de posse destas informações, determinar qual a estratégia a seguir para atender às diferentes necessidades dos clientes identificados. Normalmente utiliza recursos de mineração de dados para localizar padrões de diferenciação entre os clientes.
- CRM Colaborativo: é a aplicação da tecnologia de informação que permite a automação e a integração entre todos os pontos de contato do cliente com a empresa. É importante preparar esses pontos de contato para interagir com o cliente e disseminar as informações levantadas para os sistemas do CRM Operacional.
- CRM Social: É a forma de interagir com o cliente por meio das mídias sociais, e ainda de enriquecer os dados e informações sobre o cliente com base nas informações encontradas em seus perfis nas redes sociais.

Uma empresa pode utilizar uma ou mais aplicações dependendo do tipo de negócio, perfil do cliente a ser abordado e tamanho da empresa. Todas as aplicações podem ser trabalhadas de maneira integrada e complementar a fim de entregar o melhor resultado.

De acordo com Judy Strauss e Raymond Frost (2011), a implantação de uma solução de CRM em uma empresa é responsável por diversos benefícios, como redução de custos, aumento de receita e retenção de clientes. Tudo isso é resultado de estudo das informações armazenadas no banco de dados, que ajudam a empresa a compreender melhor seus clientes e com isso promover a lealdade através da satisfação do cliente. Clientes leais custam menos e garantem um retorno positivo para a empresa.

Segundo Kotler (1998), é muito mais barato manter os clientes já contidos na carteira do que buscar novos clientes no mercado. Além disso, as equipes de venda se tornam mais eficazes quando conhecem melhor os clientes. Consequentemente, clientes satisfeitos além de comprarem mais, fazem recomendações positivas e essas recomendações são de extrema valia para a empresa, pois essa comunicação boca a boca pode atrair muitos outros clientes.

Atualmente, quando se fala em comunicação boca a boca, não se trata apenas de divulgação pessoal, essa divulgação costuma ser feita através de redes sociais, blogs, aplicativos de comunicação e muito mais, por isso é importante que as empresas fiquem atentas a esses novos canais de comunicação e estejam preparadas para interagir com os clientes através deles.

Segundo Judy Strauss e Raymond Frost (2011), as vantagens econômicas e estratégicas de um sistema de CRM são:

- Aumento do percentual de fidelização dos clientes.
- Economia de tempo e aumento da produtividade com automatização de tarefas.
- Otimização e integração entre os diversos serviços da empresa como Comercial, *Marketing* e Serviço pós-venda.
- Contribuição para a vantagem competitiva da empresa.
- Aumento dos lucros da empresa.

Antes de se tomar a decisão de implantar pacotes de *software* CRM, é importante que a empresa analise se a base de dados e/ou *data warehouse* são suficientes para a solução desejada. Para isso, é necessário consultar os técnicos responsáveis pelas bases de dados, avaliar o ambiente de hardware, de rede e *software*. O resultado dessa consulta será a priorização de solução de problemas que possam impedir a implantação.

3.2 Sistema Bancário Brasileiro

De acordo com o livro História do Banco do Brasil (2010), o sistema bancário brasileiro se iniciou em 1808 quando o príncipe D. João, recém chegado ao Brasil, decidiu criar o Banco do Brasil.

Com o crescimento da exportação do café, foram surgindo novas instituições financeiras, à princípio firmas comerciais que procuravam funcionar como casas bancárias em limitadas operações de câmbio. Até que em 1838, foi fundado o Banco Comercial do Rio de Janeiro.

Em 1857 houve uma reforma bancária por determinação do ministro da fazenda Souza Franco, que afirmava ser necessária a criação de bancos em outros estados do país para facilitar o uso dos capitais por prazos mais longos e melhor atender aos interesses dos depositantes de fundos impulsionando o crescimento da produção, tão limitado no Império.

De acordo com Chavantes (2003), foram fundadas em 1920 a Inspeção Geral dos Bancos, em 1921 a Câmara de Compensação do Rio de Janeiro e em 1932 a Câmara de Compensação de São Paulo.

Em 1929 houve uma grande crise, conhecida como A crise de 1929 ou de A Grande Depressão. Essa crise começou nos Estados Unidos que exportavam itens industrializados e agrícolas para a Europa. Em 1929, as cidades europeias já estavam recuperadas da Primeira Guerra Mundial e decidiram reduzir significativamente a importação de produtos dos Estados Unidos.

Com isso, as indústrias dos Estados Unidos passaram a acumular muitos produtos em estoque. Grande parte dessas indústrias possuía ações na Bolsa de

Valores de Nova York e muitos americanos investiam nessas ações. Com a redução da exportação, as ações das empresas começaram a se desvalorizar e os investidores imediatamente começaram a tentar vender suas ações. O efeito foi devastador e muitas pessoas da classe alta ficaram pobres na mesma velocidade em que as ações se desvalorizavam.

Como na época os Estados Unidos possuíam parcerias comerciais com muitos países, a crise acabou se espalhando pelo mundo.

No Brasil o efeito foi sentido com relação à redução da exportação do café, porque na época, os Estados Unidos eram os maiores compradores de café do Brasil. O preço do café diminuiu muito e para que não houvesse uma desvalorização excessiva, o governo brasileiro comprou e queimou toneladas de café. Fazendo isso, foi possível reduzir a oferta e manter o preço desse importante produto para a economia da época. Essa crise também teve um efeito positivo para o Brasil, porque com a diminuição da procura, muitos empresários do café começaram a investir em outros setores, como o industrial e conseqüentemente alavancaram a indústria brasileira.

Após a Segunda Guerra Mundial, novas instituições financeiras foram criadas, como o FMI (Fundo Monetário Internacional) e o Banco Mundial. Em 1945 a Superintendência da Moeda e do Crédito (SUMOC), que em 1964 se tornou o Banco Central do Brasil. O BNDES, o Sistema Financeiro da Habitação, o Banco Nacional da Habitação e o Conselho Monetário Nacional também foram criados entre as décadas de 50 e 60. O país passa a ser regulamentado através do Conselho Monetário Nacional (CMN) e do Banco Central (BC, BaCen ou BCB), que se transformam nos principais órgãos do sistema financeiro brasileiro.

No final de 1976 foi criada a Comissão de Valores Mobiliários (CVM), cujo objetivo é facilitar a liberação de recursos para as empresas, pois começaram a surgir bancos de investimentos e a obtenção de recursos exteriores para as empresas foi facilitada pelo CVM. O Sistema Especial de Liquidação e Custódia (SELIC) foi criado em 1979 e o objetivo era realizar a custódia e liquidação com títulos públicos como as Letras do Tesouro Nacional e as Obrigações Reajustáveis do Tesouro Nacional.

Em 1988, a Constituição Brasileira estruturou o sistema financeiro nacional com o intuito de promover o desenvolvimento e equilíbrio do país, a estabilidade econômica e os interesses de toda a população. Os bancos tornaram-se múltiplos,

ou seja, não estariam limitados apenas ao setor financeiro, podendo atuar em outros ramos da economia. Surgiram produtos como a previdência privada que passaram a ganhar maior atenção do público.

De acordo com Luiz Carlos Bresser Pereira (1994), entre dezembro de 1993 e julho de 1994 foi implantada uma reforma monetária no país que transformou a Unidade Real de Valor (URV), o Cruzeiro Real em Real. O Plano Real foi implantado em 3 fases, com a primeira se iniciando em dezembro de 1993. Dentre os 13 planos de estabilização tentados no Brasil, esse foi o mais bem-sucedido porque adotou uma solução rigorosa inovadora: a coordenação prévia dos preços relativos através da URV. Com isso foi possível neutralizar as defasagens nos aumentos de preços que caracterizam a inflação crônica ou inercial.

O Plano Real adotou uma abordagem rigorosamente de mercado, evitando congelamentos. Todavia, foi realista ao tornar obrigatória a conversão dos salários para a URV no início de sua segunda fase. Dessa forma, o problema de indexação salarial foi contornado.

Após a criação do Plano Real e estabilização da economia, aconteceu a popularização dos cartões de crédito. Esse tipo de cartão já existia no Brasil desde a década de 1950, eram cartões *Diners Club* cuja emissão era feita por meio de franquia e inicialmente eram conhecidos como cartões de compras. Em 1968 foi lançado o primeiro cartão de crédito por um banco. O cartão de débito foi lançado bem depois, somente em 1983.

A partir de 1995, os cartões de débito e crédito principalmente, começaram a ganhar bastante espaço no mercado tornando-se um dos principais meios de pagamento utilizados até os dias atuais.

Nesse mesmo período, os empréstimos pessoais tornaram-se mais populares também. O objetivo era fomentar a aquisição de bens de consumo duráveis, por meio de financiamentos ou consumo de bens não duráveis por meio de crédito pessoal ou limite de crédito em conta corrente ou cartão de crédito.

Além dos bancos, lojas de grande porte começaram a oferecer cartões de crédito próprios ou em parceria com grandes administradoras de cartões de crédito. Todo esse processo ampliou as possibilidades desse novo mercado e tornou possível compartilhar dados em tempo muito curto ou real entre as instituições envolvidas. São Paulo sempre foi destaque em termos de eficiência quanto às

necessidades do setor e se consolidou como o grande concentrador dos conglomerados financeiro nacional.

Uma grande crise econômica Global se iniciou em 2008 e é sentida até os dias de hoje. A Grande Recessão como ficou conhecida, teve sua origem com a falência do tradicional banco de investimento Lehman Brothers, fundado em 1850. Outras grandes empresas quebraram em seguida e esse processo ficou conhecido como “crise dos subprimes”.

O PIB da Zona do Euro teve uma queda de 1,5% no quarto trimestre de 2008, em relação ao trimestre anterior, a maior contração da história da economia da zona do Euro.

O Brasil foi afetado pela crise por conta da globalização dos negócios entre países. O efeito mais imediato foi a baixa das cotações das ações em bolsas de valores, por causa do aumento expressivo das vendas de ações de especuladores estrangeiros. Por conta disso, também houve uma rápida alta do dólar. Com as empresas brasileira exportadoras vendendo menos, a recessão atingiu o país e postos de trabalho começaram a ser cortados.

No sistema bancário, houve uma retração dos empréstimos concedidos pelos grandes bancos aos menores. Mesmo com a isenção de uma parte do depósito compulsório cedido pelo Banco Central do Brasil, os grandes bancos decidiram não conceder os empréstimos devido ao clima quase de pânico que se instaurou nos mercados financeiros em geral.

Atualmente o sistema bancário brasileiro enfrenta os mais diversos desafios, muitos deles em decorrência da modernização, expansão, internacionalização e evolução tecnológica. Por serem estruturas de grande porte, burocráticas e com diversos órgãos regulamentadores, os bancos têm perdido mercado para as atuais *Fintechs*.

Segundo Fraga (apud Venture Scanner, 2016), existem pelo menos 1.406 *Fintechs* no mundo. *Fintech* é o nome dado a empresas *startups* que fazem exatamente o que um banco faz. A diferença é que a estrutura de uma *fintech* é mais enxuta, o que proporciona redução de custo dos serviços para os clientes, agilidade e o alcance ilimitado da internet.

Fintechs mudam a forma como as pessoas pagam compras, administram contas, investem dinheiro e contratam empréstimos. É possível por exemplo abrir

uma conta via aplicativo de celular e receber o cartão em casa e contratar cartão de crédito sem taxas.

O sistema bancário aposta em parcerias com essas empresas, dessa forma, tem uma oportunidade de não ficar para trás e ao mesmo tempo as *Fintechs* validam seus negócios, ganham experiência e recebem investimentos.

Para isso, o uso do *Data Science* para conhecer melhor o perfil dos seus clientes através do CRM é fundamental para embasar as mudanças de produto, serviço e processo, bem como, a criação de novos produtos e serviços.

3.3 O uso de CRM no Ambiente Bancário

Segundo Berger et al (2009), uma implementação competitiva de CRM em bancos necessita de um processo de negócio alinhado com a contribuição de valor.

Joe Peppard (2000), sugere que as instituições financeiras em particular, devem adotar uma perspectiva de toda a empresa, envolvendo desde os colaboradores que têm contato direto com os clientes finais até os que trabalham na operação do negócio sem contato direto com os clientes. Ele acredita que, dessa forma, a instituição financeira se tornará verdadeiramente “centrada no cliente” e conseguirá capitalizar sobre as oportunidades oferecidas pelas informações que essa atitude trará.

O processo de CRM de um grande banco brasileiro foi analisado e nessa instituição existe uma área de negócios específica de CRM que trata dos assuntos de forma centralizada, exceto Fraudes e Crédito. A decisão de não incluir Fraudes e Crédito no processo de CRM centralizado foi tomada por conta da fragilidade e legislação rigorosa que envolve os dois tópicos. Sendo assim, todos os assuntos descritos com relação à área de CRM desse banco, não considerará Fraudes e Crédito.

A área de CRM do banco possui acesso a informações internas (produtos contratados, dados cadastrais, histórico de atuação...) e externas (bases cadastrais compradas, Receita Federal, parceria com outras empresas...) e com base nesses acessos é possível ter uma visão mais completa dos clientes, facilitando estudos de perfis e comportamentos.

Joe Peppard (2000), afirma que em um número crescente de organizações, as iniciativas de CRM estão sendo conduzidas pelo departamento de *Marketing* que visualiza o resultado final de um repositório de informações do cliente e busca usar técnicas de *Data Mining* para extrair padrões através desses dados para fins de tomada de decisão de *Marketing*.

A princípio, a equipe de CRM desse grande banco fazia parte do departamento de *Marketing*, sendo que, com o passar do tempo eles adquiriram experiência e tomaram a decisão de separar as áreas. Atualmente, a área de CRM trabalha em parceria com a área de *Marketing* que é responsável por focar em assuntos especificamente de *Marketing*, como por exemplo, a melhor forma de apresentar uma campanha na tela do internet banking do cliente.

Segundo Joe Peppard (2000), instituições de serviços financeiros européias são capazes de entregar valor através do que ele denomina investimentos ECRM, mas só o fazem porque possuem uma ampla perspectiva desse conceito.

Ainda segundo Joe Peppard (2000), essa perspectiva pode ser descrita através do acrônimo ECRM, cujas letras descrevem os quatro conceitos centrais:

E – *E-business: enterprise business* ou negócio corporativo, o mais importante é a integração das atividades do *e-business* com o *framework* de todas as atividades comerciais existentes e futuras.

C – *Channel management*: Canal de gestão. O canal do melhor impacto ou economia a qualquer momento, a qualquer lugar e para todos. Canais de acesso e distribuição integrados e interativos.

R – *Relationships*: Relacionamentos comerciais reais construídos em serviços de excelência, valor e conveniência.

M – *Management*: Gerenciamento de toda a empresa. Integração total dos processos de *back-office* e *front-office*.

Abaixo, segue a descrição do processo, cotidiano, principais gaps e evoluções do CRM do grande banco brasileiro, na visão de um especialista integrante da equipe não-técnica do time de CRM:

Ele descreve a área de CRM como um grande usuário de dados de toda a empresa. Porque a área tem acesso a quase todas as informações de todos os

segmentos do banco, como conta corrente, cartão de crédito, financiamentos (veículos, imóveis, etc), seguros, etc.

Essas informações não estão centralizadas em um único repositório de dados e podem ser capturadas de diversas fontes e formatos. Isso se torna um problema por conta do tempo gasto para localizar e interpretar as informações.

O objetivo da área de CRM desse banco é, basicamente, estudar o perfil dos clientes, para manter o relacionamento, ofertar o produto e/ou serviço adequado, através do canal e momento adequado para o cliente certo, afim de trazer receita para o banco e evitar a perda de cliente para a concorrência.

O ato de fazer contato para manter o relacionamento e/ou ofertar produtos e serviços para os clientes é chamado de campanha.

De acordo com o especialista integrante da equipe não-técnica do time de CRM, a cadeia de processos de criação de uma campanha de CRM no grande banco brasileiro citado é composta por:

- **Necessidade:** pode ser uma observação identificada pela área de Produtos ou uma adaptação necessária por conta das ações que o mercado executa. Por exemplo, pode ser necessário aumentar a taxa de venda de um produto de conta corrente para clientes universitários.
- **Dados:** componente principal do processo de CRM. Sem dados é impossível iniciar uma campanha.
- **Estudos:** são análises de dados, cujo objetivo é identificar perfis de clientes. Esses estudos se baseiam em uma necessidade, que basicamente responde uma pergunta, como por exemplo: Qual o perfil de cliente mais propenso a comprar o produto para universitários?
- **Modelos:** é o resultado do estudo realizado, onde identifica-se o público alvo para a campanha. No modelo é possível identificar as principais características dos clientes como segmentação financeira, renda, etc.
- **Teste:** os modelos são testados, ou seja, a campanha é ofertada para um percentual pequeno dos clientes do modelo para validar sua efetividade. Caso a aderência esteja em um nível considerado aceitável (esse nível pode variar

de acordo com o objetivo da campanha, tipo de produto, segmento e tipo de cliente), a campanha é implantada.

- Implantação: consiste em divulgar a campanha para todo o público do modelo.

De acordo com o mesmo especialista, as campanhas são divulgadas para clientes específicos e por diversos canais de comunicação. Para identificar o melhor cliente e canal. As três perguntas abaixo precisam ser respondidas:

- Quem é o cliente?
- Qual o melhor canal para esse cliente?
- Qual o melhor momento de oferta para esse cliente?

Encontrar as respostas para essas perguntas, atualmente, não é uma tarefa fácil porque na maior parte das vezes, esse trabalho é feito através do cruzamento de dados que podem estar em planilhas Excel e bases Access por exemplo. Tanto a construção quanto a atualização é feita manualmente e isso causa muita dificuldade porque o tempo gasto com essas atividades pode comprometer a oferta do produto ou serviço.

O índice de assertividade conseguido também não costuma ser muito alto, o que impacta diretamente a campanha porque, consequentemente, a oferta não é feita para o público mais adequado. Isso acarreta em perda financeira devido aos gastos com a divulgação da campanha e desgaste de imagem junto ao cliente final, já que o mesmo é incomodado com o recebimento de ofertas que não estão de acordo com seu estilo de consumo.

Uma campanha pode ser ofertada para um cliente através de um ou mais canais de comunicação, porém, é necessário garantir que após a contratação do produto ou serviço, o cliente não receberá mais a oferta. Dependendo das características da campanha, esse trabalho precisa ser feito de forma online. Por exemplo: se uma campanha está ofertando um produto via internet e telefone e o cliente aceita contratar o produto via internet, é importante garantir que esse cliente não receberá uma ligação ofertando o produto que ele acabou de contratar.

O tempo é um fator muito importante para o sucesso de uma campanha de CRM e o tempo de resposta especificamente desse banco costuma comprometer a

efetividade das campanhas. Isso acontece principalmente pelo tamanho da estrutura e processos complexos a que as atividades precisam ser submetidas para serem concluídas.

Dependendo da demora para responder à solicitação de um cliente, o banco corre o risco de perdê-lo para o concorrente de mercado mais ágil e isso também é um problema que demanda muito esforço.

Na visão técnica de um especialista que trabalha na área de Tecnologia de CRM do mesmo grande banco brasileiro, o principal papel da Tecnologia no processo de CRM do banco é automatizar a divulgação da campanha da melhor forma possível, de acordo com as variáveis que melhor atendem o cliente final. O trabalho da equipe de Tecnologia é executado através de metodologias ágeis, o que acelera a divulgação da campanha para o público alvo.

Como o modelo já chega pronto para a equipe de Tecnologia, o foco do trabalho é realmente estudar a melhor maneira de divulgar a campanha.

Eventualmente a equipe de Tecnologia também auxilia na validação dos modelos propostos através do uso de ferramentas e metodologias que analisam o comportamento desses clientes, inferindo um padrão de comportamento com o intuito de complementar o modelo aumentando a taxa de assertividade.

As validações dos modelos são feitas com base nas informações disponíveis em uma base histórica de dados não-estruturados.

A base histórica é usada também para a construção de motores.

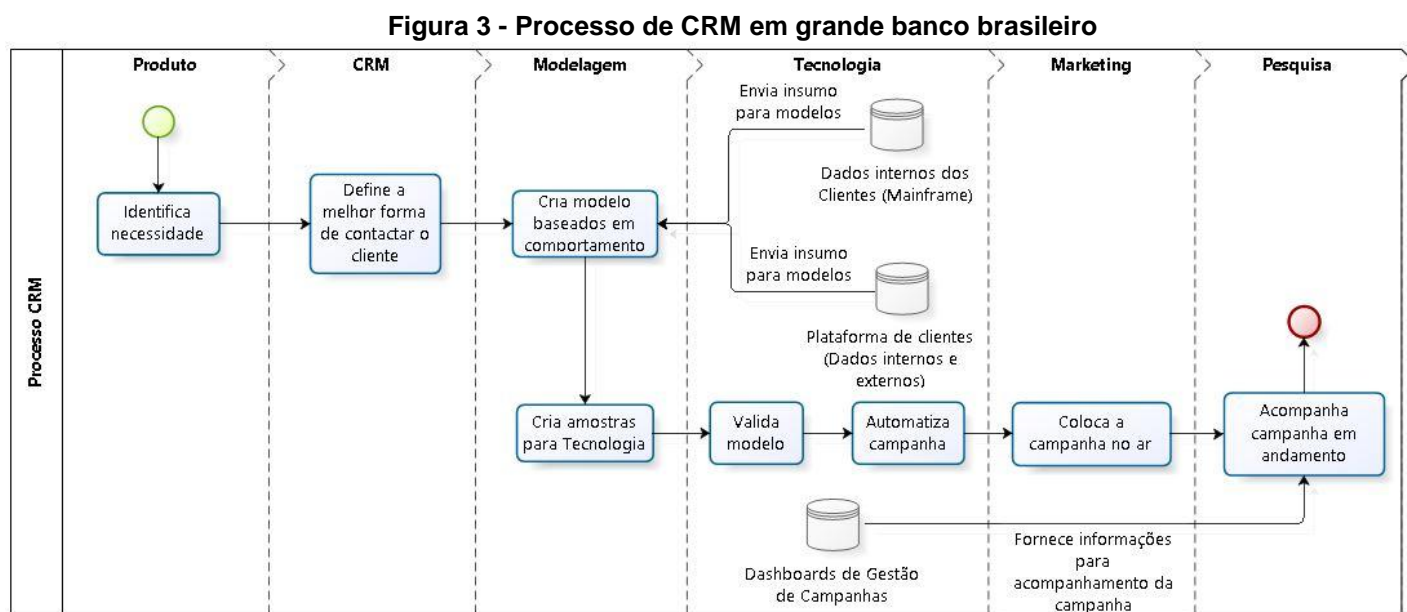
Os motores são responsáveis por interpretar características e criar indicadores do potencial de compra do cliente. Esses indicadores são disponibilizados em uma plataforma para a área de CRM e Modelagem e agilizam a construção de perfil de clientes.

Baseados nesses indicadores, a área de Modelagem também é capaz de criar amostras de clientes que são usadas para validar um determinado modelo que futuramente será usado em uma campanha.

Para acompanhamento de campanhas que estão em execução, existe um sistema de gestão centralizador de campanhas, onde é possível fazer o acompanhamento através de *Dashboards*. O acompanhamento é diário e o sistema faz a captura de dados de todos os canais em que as campanhas têm acesso de forma automática. O objetivo desse sistema de gestão é avaliar a efetividade da

campanha e tomar decisões como, por exemplo: retirar a campanha do ar, efetuar ajustes, aumentar o prazo da campanha, etc.

Com base nas descrições colhidas dos especialistas técnicos e funcionais, pode-se dizer que o processo de CRM do banco se comporta conforme apresentado na figura 3):



Fonte: A autora

A necessidade de criação de uma campanha é identificada pela área de Produtos que faz a solicitação para a área de CRM.

A área de CRM estuda a necessidade enviada por Produtos e define a melhor forma e canal em que a campanha deve ser divulgada, com base nas características do cliente.

A área de Modelagem é responsável por criar o modelo que será usado na campanha e eventualmente a base amostral para que a área de tecnologia consiga validar o modelo criado.

Na área de Tecnologia estão as bases que são usadas como insumo tanto para a criação dos Modelos quanto para o acompanhamento do andamento das campanhas por parte da área de Pesquisa de resultados.

Por fim, a área de *Marketing* é responsável por criar a identidade visual da campanha e colocá-la no ar, levando em consideração todas as premissas técnicas e de negócio que envolvem o produto em questão.

3.4 Considerações de CRM e o Sistema Bancário

O CRM é uma metodologia de gestão que auxilia as empresas de qualquer ramo a melhorar o relacionamento com o cliente final, através de ferramentas e análises.

Antes de se tomar a decisão de usar o CRM para alavancar vendas, aumentar a retenção ou simplesmente entender o ponto de vista do cliente, é necessário realizar um estudo interno para definir quais são os objetivos e se necessário adaptar processos, metodologias para que a cultura da empresa inteira esteja voltada para o bom atendimento do cliente final.

A tecnologia tem um papel significativo nesse processo, pois é responsável pela integração dos sistemas, captura, organização, processamento e auxílio na interpretação dos dados. A tecnologia é importante também por conta do tempo de resposta necessário para atender necessidades específicas de cada cliente.

O sistema bancário brasileiro tem sofrido grandes e significativas mudanças desde sua criação, principalmente a partir do momento em que empréstimos pessoais, cartões de crédito e débito tornaram-se mais acessíveis e populares.

A situação tornou-se ainda mais relevante com o aumento de *fintechs* que prestam o mesmo tipo de serviços que os bancos e oferecem os mesmos produtos, porém numa velocidade muito maior, menos burocrática e sem taxas administrativas.

Os bancos precisam cada vez mais estudar o perfil dos clientes para entender suas necessidades e atendê-las de maneira que consigam evitar a migração de clientes para esses novos tipos de empresa.

Analisando o processo de CRM de um grande banco brasileiro, foi possível identificar oportunidades de melhoria em pontos já conhecidos por quem estuda o assunto. O tempo de resposta e dificuldade de implementar ajustes em uma

campanha costumam ser pontos de atenção mencionados por todos os envolvidos no processo de criação e divulgação de campanhas CRM.

Além disso, é necessário evoluir o processo de criação de modelos, com a redução da pulverização de dados e aumento na qualidade das bases históricas.

Outro ponto mencionado como importante, é a integração das análises geradas pela equipe de Pesquisa de resultados com as demais áreas envolvidas no processo. É sabido que o percentual de assertividade das campanhas pode melhorar desde que se tenha insumo suficiente para identificar os pontos de melhoria, mas como a maior parte das informações e do conhecimento necessário para se gerar essa análise fica retido na área de Pesquisas, a assertividade das campanhas evoluem numa velocidade abaixo da esperada pelo banco.

4. ANÁLISE DOS RESULTADOS

Com base em análises de tendência de mercado e indicadores de resultado das campanhas colocadas no ar, o banco mencionado no capítulo anterior identificou que suas campanhas em geral não estavam obtendo o resultado esperado e que o nível médio de efetividade estava caindo.

Ciente das tendências do mercado que está cada vez mais dinâmico com o advento das *fintechs* e tecnologias em geral e levando em consideração o tempo de resposta comprometido por conta do tamanho de sua estrutura, o banco optou pela contratação de uma consultoria americana, cujo objetivo era analisar o cenário atual do banco e fazer uma proposta que alavacasse as vendas de produtos e serviços e melhorasse o relacionamento do banco com os clientes.

A consultoria em questão é especializada desenvolver soluções que auxiliam agências de inteligência do governo americano, setor militar e justiça a integrarem e analisarem seus dados para o combate do terrorismo e do crime. Desde 2010, se especializou também em soluções para clientes da área financeira.

A equipe da consultoria identificou que se os modelos utilizados para divulgação das campanhas fossem melhorados, o retorno das campanhas poderia ter um aumento acima de 25%. Esse percentual foi comprovado através de um dos modelos construídos na prova de conceito realizada.

4.1 Prova de conceito para cases de campanhas de CRM

Para realizar a prova de conceito, foi necessário construir um ambiente apartado.

A infraestrutura desse ambiente era composta por um servidor de aplicação e processamento dedicado. Uma rede de alto desempenho de 10 GB com um canal alternativo de 1GB e sala de trabalho exclusiva com acesso remoto via VPN e dados sem mascaramento.

As ferramentas utilizadas em sua maioria era de propriedade da consultoria em um ambiente chamado Foundry.

O Foundry é composto por 4 camadas:

A primeira camada é a camada de interface de consulta, onde se utiliza uma linguagem proprietária chamada MESA, que é usada para abstrair aos cientistas de dados a complexidade de outras linguagens de programação.

A segunda camada é a camada de software que administra os dados existentes no cluster, bem como as bases criadas pelos usuários do ambiente e suas dependências.

A terceira camada é a de *framework* de *cluster computing*, compatível com Hadoop onde foi usado Spark, que é um framework para processamento de Big Data.

E a quarta e última camada é a camada de plataforma de computação distribuída onde foi usado Hadoop de fato.

Como ferramenta de apresentação de informações consolidadas, foi utilizado Slate, ferramenta proprietária que buscou as informações a partir de banco de dados relacional PostgreSQL.

Para desenvolvimento web, foi utilizada a linguagem proprietária Tropic of Cancer.

E finalmente para visualização de alto volume de dados direto no Hadoop, utilizou-se o Contour, também proprietário da consultoria.

Mais de 280 bases de dados foram carregadas no ambiente criado e após o carregamento foi realizado um processo de integração dos dados através da seleção de variáveis que compuseram uma base única com todas as informações relevantes para o desenvolvimento das análises.

Na base única, foram aplicadas as técnicas de Data Science descritas no capítulo 2 em conjunto com as ferramentas proprietárias da consultoria.

4.1.2 Modelo de gasto de cartão de crédito

Um dos modelos construídos pela consultoria, tinha como objetivo identificar oportunidades de aumento de gastos com cartão de crédito dos clientes porque o modelo atual não é otimizado para as necessidade do cliente.

Através da base única criada foi possível ter uma visão integrada de toda a carteira de cartão de crédito e a capacidade de interagir, construir e testar modelos robustos de ciclo de vida do cliente que permitiu a identificação de:

- Variáveis preditivas que levam ao aumento de gasto
- Períodos de tempo, segmentos de clientes e outras variáveis que indicam atraentes oportunidades de venda cruzada

Identificação de alavancas estratégicas onde essas percepções podem ser operacionalizadas.

Plano de desenvolvimento ou protótipo de ferramentas operacionais para unidades de negócios, por exemplo, atendimento ao cliente ou *Marketing* para aproveitar melhor essas ideias em operações de negócio.

O primeiro passo foi identificar o produto certo para o cliente certo através do canal certo.

Para essa atividade foram utilizadas 30 fontes de dados, dentre elas, dados demográficos do cliente, produtos contratados, transações de débito e crédito, vendas via canais, pagamentos, etc.

O resultado desses cruzamentos foram 8 modelos de propensão de produtos, 8 modelos de receitas esperadas com 19 mil características.

Uma amostra de 15 mil clientes foi utilizada para testar o modelo criado pela consultoria e o modelo atual do banco.

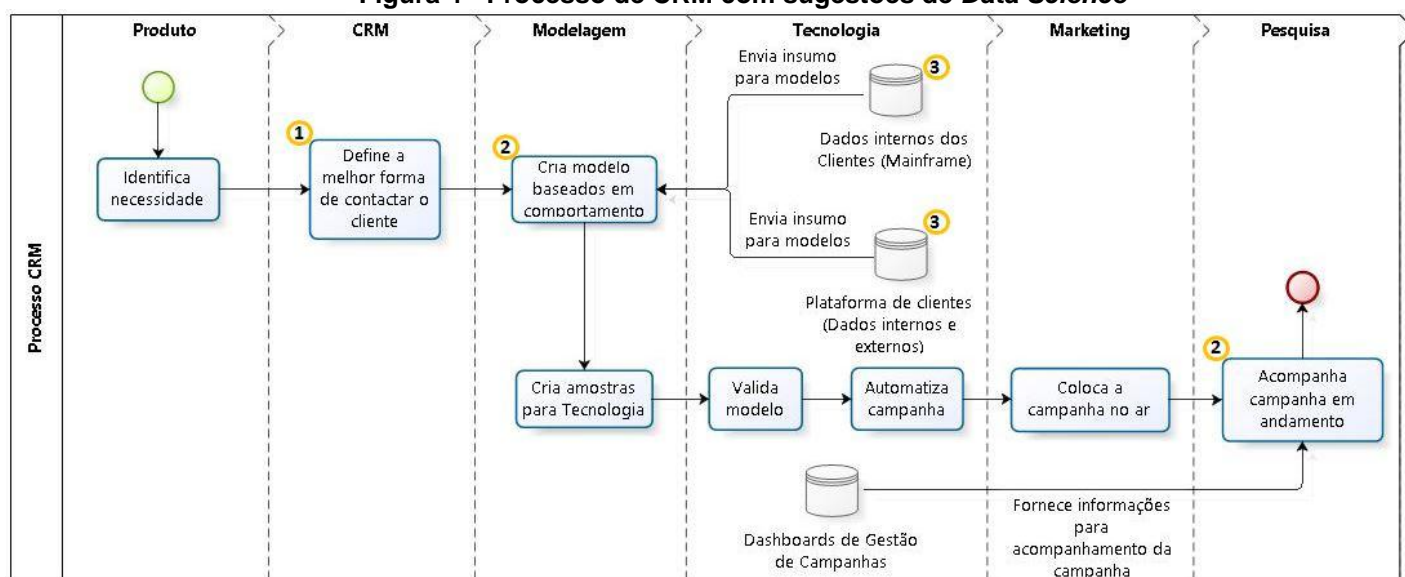
O modelo da consultoria resultou em um retorno 25,60% maior que o modelo atual do banco.

4.2 A aplicação do *Data Science* para CRM no Ambiente Bancário

Analisando o processo de CRM descrito na Figura 3 e levando em consideração as técnicas estudadas no capítulo 2, é possível identificar pontos de melhorias além dos que os especialistas técnicos e funcionais mencionaram.

A Figura 4, contém marcações nos pontos que podem ser beneficiados por técnicas de *Data Science* e seus detalhamentos.

Figura 4 - Processo de CRM com sugestões de *Data Science*



Fonte: A autora

4.2.1 Análises estatísticas

O ponto sinalizado com o número 1 na Figura 4, é a atividade onde a equipe de CRM define a melhor forma de contactar o cliente. Essa atividade é realizada de forma manual, utilizando técnicas não revisadas com frequência, o que a torna menos eficiente.

A proposta de se utilizar análises estatísticas, refere-se à utilização ativa das técnicas, sempre levando em consideração as premissas apresentadas pela área de produtos.

Com a estatística descritiva, é possível organizar, descrever e sumarizar os dados que precisam ser analisados.

Já com a probabilidade é possível identificar relações aleatórias entre os dados, como por exemplo, a quantidade de pessoas que contratam determinado serviço após um horário específico.

E por meio da inferência, é possível interpretar o que aquela determinada informação significa e como pode ser usada para agregar valor ao negócio.

Através da prova de conceito criada pela consultoria americana, comprovou-se numericamente que com as técnicas estatísticas é possível definir com mais segurança quais variáveis são relevantes para cada cliente específico.

Também é possível criar grupos segmentados de acordo com a identificação de comportamentos em comum.

4.2.2 Técnicas de Machine Learning

Os pontos sinalizados com o número 2 na Figura 4 são os pontos onde as técnicas de *Machine Learning* foram testadas.

A atividade da equipe de Modelagem, onde os modelos são criados baseados no estudo do comportamento dos clientes e a atividade da área de Pesquisa, onde são feitos os acompanhamentos de retorno das campanhas seriam beneficiadas pelas técnicas de *Machine Learning* porque com isso é possível ensinar o computador a interpretar os resultados apresentados e consequentemente apresentar análises cada vez mais profundas e assertivas.

Através da prova de conceito criada pela consultoria americana, a técnica de Aprendizagem Supervisionada foi aplicada na atividade da equipe de Modelagem. O objetivo era melhorar a busca por padrões na base pré-definida pela equipe de CRM.

Constatou-se que a aplicação da técnica realmente aumentou a precisão dos padrões encontrados comparando-os com os padrões encontrados através do método utilizado atualmente.

Com relação à atividade da equipe de Resultados, foram aplicadas técnicas de Classificação e Regressão para interpretar os resultados de uma forma mais consistente e clara. Atualmente, não existe validação de resultados de campanha por parte da área de CRM utilizando as informações geradas pela área de Resultados, sendo assim, não é possível fazer comparações.

Outras técnicas de *Machine Learning* foram aplicadas para as atividades em questão, os exemplos citados servem apenas de apoio para comprovar a eficácia das técnicas.

4.2.3 Técnicas de Data Mining

Os pontos sinalizados com o número 3 na Figura 4 referem-se às bases de dados que são as principais fontes para criação de modelos de campanhas.

Atualmente as fontes de dados do banco são grandes bases de dados, distribuídas de forma heterogenia. Esse cenário que é muito comum atualmente e exige a aplicação de técnicas e ferramentas específicas para que se possa extrair conhecimento da massa de dados.

No banco em questão, a base Plataforma de clientes foi construída com o propósito de centralizar o máximo de informações de clientes disponível e com o auxílio de ferramentas de *Data Mining* são aplicadas técnicas para oferecer informações que ajudam na compreensão dos dados e criação dos modelos de perfil de clientes.

A flexibilidade de se trabalhar tanto com dados internos quanto externos dos clientes é uma grande vantagem no processo de conhecimento e fonte de ideias (*insights*) até então desconhecidas. Tendo conhecimento e analisando o comportamento dos clientes dentro e fora do banco, é possível descobrir informações que enriquecem o modelo de forma significativa.

4.3 Considerações das Análises de Resultado

A prova de conceito motivada pela necessidade de melhoria do retorno das campanhas criadas pelo banco além de comprovar numericamente sua eficácia, também contribuiu para modernizar a forma de pensar das pessoas envolvidas nas atividades de criação de campanhas.

As novas ferramentas e técnicas utilizadas apresentaram um resultado acima do esperado. Essa constatação potencializou a motivação para mudança e auxiliou na redução da resistência natural que se apresenta quando uma mudança muito grande é proposta.

A criação da Plataforma de clientes com dados não estruturados e uso de ferramentas específicas de *Data Mining* já apresenta resultados no processo.

5. CONSIDERAÇÕES FINAIS

Levando em consideração o cenário atual, é possível dizer com segurança que as empresas dependem cada vez mais de recursos tecnológicos que acompanham as tendências de mercado e consequentemente refinam o gosto dos clientes tornando-os mais exigentes.

Essa hipótese pode ser comprovada através da frequência cada vez maior de números estagnados ou em declínio, nos relatórios de resultado de campanhas de um grande banco brasileiro.

Ao analisar esses números detalhadamente, foi possível identificar o crescimento de um novo nicho de mercado, as *fintechs*, e entender que os concorrentes atuais também estavam executando ações de modernização tecnológica para conseguir novos clientes e fidelizar os clientes atuais.

Com os clientes cada vez mais exigentes e desprendidos de rotinas tradicionais, os bancos em geral, se viram obrigados a traçar uma estratégia para acompanhar as tendências do mercado.

E uma boa forma de identificar quais ações tem efeito mais eficaz, é conhecendo o perfil dos clientes. A área de CRM, cujo objetivo é manter o bom relacionamento com os clientes é fundamental para esse trabalho.

As técnicas do *Data Science* ajudam nessa tarefa por conta da grande quantidade e variedade de dados não estruturados que são armazenados nas empresas atualmente.

Com base em uma prova de conceito realizada, foi possível comprovar a eficácia das técnicas do *Data Science* e analisando as atividades da área de CRM, sugestões de melhorias foram apontadas em partes específicas do fluxo com o intuito de aumentar o nível de assertividade das campanhas criadas.

REFERÊNCIAS

AGARWAL, SHIVAM. Data Mining Concepts and Techniques. Disponível em: <<https://www.deepdyve.com/lp/institute-of-electrical-and-electronics-engineers/data-mining-data-mining-concepts-and-techniques-THuMpNCKpt>>. Acesso em: 06 de junho de 2016.

BERGER, S. C.; GENSLER, S.; SKIERA, B.; WIESEL, T. Customer *Relationship Management* Processes in Retail Banking – Best Practice and Business Value Evaluation. Disponível em: <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4755679&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D4755679>. Acesso em: 22 de abril de 2016.

BERRY, JONATHAN. Data Base *Marketing*. Disponível em: <<http://www.bloomberg.com/news/articles/1994-09-04/database-marketing>>. Acesso em: 02 de junho de 2016.

BOYER, CARL B. História da matemática. São Paulo: Edgard Blücher. 1996.

COBRA, MARCOS. *Marketing* Básico: uma perspectiva brasileira. São Paulo: Atlas. 2009.

BORNE, K. D.; JACOBY, S.; CARNEY, K.; CONNOLLY, C.; EASTMAN, T.; RADDICK, M. J.; TYSON, J.A.; WALLIN, J. The Revolution in Astronomy Education: Data Science for the Masses. Disponível em: <<https://arxiv.org/pdf/0909.3895.pdf>>. Acesso em: 06 de junho de 2016.

CHAVANTES, ANA PAULA. A regulação do sistema bancário e o desenvolvimento econômico paulista, 1889-1930. Disponível em: <<http://www.revistalep.com.br/index.php/lep/article/view/17>>. Acesso em: 02 de junho de 2016.

Cleveland, S. William. *Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics*. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2001.tb00477.x/abstract>>.

Acesso em: 03 de junho de 2016.

COBRA, Marcos. *Marketing Básico: uma perspectiva brasileira*. São Paulo: Atlas, 2009.

COBRA, M.; BREZZO, R. *O novo marketing*. São Paulo: Campus, 2010.

DAVENPORT, H.T.; COHEN, D.; JACOBSON, A. *Competing on Analytics*. Disponível em: <<http://www.babsonknowledge.org/analytics.pdf>>. Acesso em: 06 de junho de 2016.

DAVENPORT, H. T.; PATIL, J. D. *Data Scientist: The Sexiest Job of the 21st Century*. Disponível em: <<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>>. Acesso em: 06 de junho de 2016.

DIRETORIA DE *MARKETING* E COMUNICAÇÃO DO BANCO DO BRASIL. *História do Banco do Brasil*. Belo Horizonte: Del Rey, Fazenda Comunicação & *Marketing*. 2010.

FAYYAD, U; SHAPIRO, G. P; SMYTH, P. *From Data Mining to Knowledge Discovery in Databases*. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>>. Acesso em: 02 de junho de 2016.

FRAGA, NAYARA. *O fenômeno Fintech: a nova leva de startups que invadiu o sistema financeiro*. Disponível em: <<http://epocanegocios.globo.com/Dinheiro/noticia/2016/01/o-fenomeno-fintech-nova-leva-de-startups-que-invadiram-o-sistema-financeiro.html>>. Acesso em: 05 de julho de 2016.

GREENBERG, PAUL. CRM, Customer *Relationship Management* na velocidade da luz: conquista e lealdade de clientes em tempo real na Internet. Rio de Janeiro: Campus. 2001.

IASC - International Association for Statistical Computing. Disponível em: <http://www.iasc-isi.org/>. Acesso em: 15 de maio de 2016.

JDS - Journal of Data Science. Disponível em: <http://www.jds-online.com/v1-1>>. Acesso em: 12 de junho de 2016.

JOHNSON, N. L.; LEONE, F. C. Statistics and experimental design in the engineering and the physical sciences. Disponível em: <https://www.ieee.org/index.html>>. Acesso em: 30 de maio de 2016.

KOTLER, P.; FOX, K. Marketing Estratégico para Instituições Educacionais. São Paulo: Atlas, 1998. 444 p.

KOTLER, PHILIP. Administração de *marketing*. São Paulo: Prentice Hal. 2005.

LAS CASAS, ALEXANDRE LUZZI. *Marketing*. São Paulo: Atlas. 2006.

Lide, R. D.; Wood, H. G. CODATA@45Years: the story of the ICSU Committee on Data for Science and Technology (CODATA) from 1966 to 2010. Disponível em: <http://www.codata.org/uploads/CODATA@45years.pdf>>. Acesso em: 12 de junho de 2016.

LOUKIDES, MIKE. What is *Data Science*. Disponível em: <https://www.oreilly.com/ideas/what-is-data-science>>. Acesso em: 02 de junho de 2016.

MONARD, M. C; BARANAUSKAS, J. A. Sistemas Inteligentes-Fundamentos e Aplicações. Barueri: Manole. 2003.

NAUR, PETER. Concise Survey of Computer Methods. Sweden: Studentlitteratur, 1974.

NAUR, PETER. The science of datalogy. Nova York: ACM, 1966.

PEDRON, C. D; SACCOL, A. Z. "What lies behind the concept of customer *relationship management*? Discussing the essence of CRM through a phenomenological approach". Disponível em: <<https://www.ieee.org/index.html>>. Acesso em: 30 de maio de 2016.

PEPPARD, JOE. Customer *Relationship Management* (CRM) in Financial Services. Disponível em: <<http://paginas.fe.up.pt/~ee94159/sne/artigo2.pdf>>. Acesso em: 14 de julho de 2016.

PEREIRA, LUIZ CARLOS B. REVISTA DE ECONOMIA POLÍTICA. A economia e a política do Plano Real. vol. 14, outubro-dezembro/94.

PLATA, SANDRA. De acordo com a vontade do cliente. Disponível em: <<http://www.intermanagers.com.br/archivo/nota.jsp?id=707>>. Acesso em: 10 de julho de 2016.

PRESSMAN, ROGER. Engenharia de *Software*: Uma abordagem Profissional. Porto Alegre: Bookman. 2011.

PROVOST, F.; FAWCETT, T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. São Paulo: O'Reilly Media, 2013. 384 p.

REDMAN, C. THOMAS. What Separates a Good Data Scientist from a Great One .Disponível em: <<https://hbr.org/2013/01/the-great-data-scientist-in-fo>>. Acesso em: 06 de junho de 2016.

SAMUEL, Arthur L. Some Studies in *Machine Learning* Using the Game of Checkers. Disponível em: < <https://www.cs.virginia.edu/~evans/greatworks/samuel1959.pdf> >. Acesso em: 17 de março de 2016.

SMITH, DAVID. "Data Science": what's in a name?. Disponível em: <<http://blog.revolutionanalytics.com/2011/05/data-science-whats-in-a-name.html>>. Acesso em: 06 de junho de 2016.

Springer Link. Disponível em: <<http://link.springer.com/journal/10618>>. Acesso em: 15 de maio de 2016.

STRAUSS, J.; FROST, R. E- MARKETING. São Paulo: Pearson; Edição: 6, 2011. 352 p.

SWAN, A.; BROWN, S. The Skills, Role & Career Structure of Data Scientists & Curators: Assessment of Current Practice & Future Needs. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.8960&rep=rep1&type=pdf>>. Acesso em: 06 de junho de 2016.

SWIFT, RONALD. CRM: *Customer Relationship Management*, o revolucionário *marketing* de relacionamento com o cliente. Rio de Janeiro: Prentice Hall. 2001.

TOZETTO, CLAUDIA; MAIA, LAURA. Startups apostam em serviços financeiros e tiram sono dos bancos. Disponível em: <<http://link.estadao.com.br/noticias/inovacao,startups-apostam-em-servicos-financeiros-e-tiram-sono-dos-bancos,10000028978>>. Acesso em: 05 de julho de 2016.

TURKEY, JOHN W. The Future of Data Analysis. Disponível em: <<http://projecteuclid.org/euclid.aoms/1177704711>>. Acesso em: 13 de fevereiro de 2016.

WITTEN, H.I.; FRANK, E. Data Mining: Practical Machine Learning Tools and Techniques. São Francisco: Morgan Kaufmann, 2005. 560 p.