



**UNIVERSIDADE ESTADUAL DO MARANHÃO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO
E SISTEMAS**

BEATRIZ NERY RODRIGUES CHAGAS

**Aplicações de Algoritmos de Aprendizado de Máquina em CRM: Revisão
Sistemática da Literatura.**

São Luís
2019



**UNIVERSIDADE ESTADUAL DO MARANHÃO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO
E SISTEMAS**

BEATRIZ NERY RODRIGUES CHAGAS

**Aplicações de Algoritmos de Aprendizado de Máquina em CRM: Revisão
Sistemática da Literatura.**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia da Computação. Centro de Ciências Tecnológicas. Universidade Estadual do Maranhão.
Área de Concentração: Ciência da Computação
Orientador: Prof^o. Ms. Antonio Fernando Lavareda Jacob Jr.

São Luís
2019

Chagas, Beatriz Nery R.

Aplicações de Algoritmos de Aprendizado de Máquina em CRM: Uma Revisão Sistemática da Literatura./ Beatriz Nery R. Chagas. – São Luís , 2020.

89 f.

Dissertação (Mestrado) – Engenharia de Computação e Sistemas, Universidade Estadual do Maranhão, 2020.

Orientador: Prof^o. Me. Antonio Fernando Lavareda Jacob Jr.

1. Aprendizado de Máquina. 2. CRM. 3. Sistemas de Informação. I. Título.

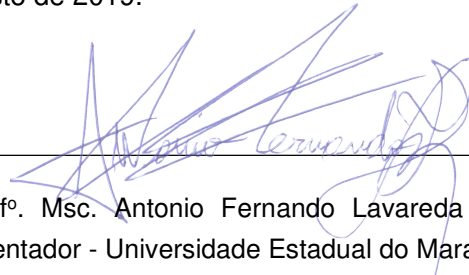
CDU: 004.42

BEATRIZ NERY RODRIGUES CHAGAS

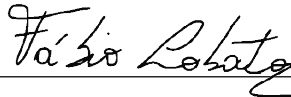
APLICAÇÕES DE ALGORITMOS DE APRENDIZADO DE MÁQUINA EM CRM:
REVISÃO SISTEMÁTICA DA LITERATURA.

Dissertação apresentada ao Programa de Pós-Graduação
em Engenharia de Computação e Sistemas da Universi-
dade Estadual do Maranhão, como parte das exigências
para a obtenção do título de Mestre em Engenharia de
Computação e Sistemas.

Aprovado em: ____ Agosto de 2019.



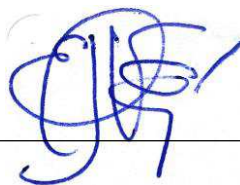
Profº. Msc. Antonio Fernando Lavareda Jacob Jr.
(Orientador - Universidade Estadual do Maranhão)



Profº. Dr. Fábio Manoel França Lobato
(Universidade Federal do Oeste do Pará)



Profº. Dr. Diego Lisboa Cardoso
(Universidade Federal do Pará)



Profª. Drª. Eveline de Jesus Viana Sá
(Instituto Federal do Maranhão)

AGRADECIMENTOS

A Deus, por me fazer forte nos momentos que pensei em desistir.

Agradeço a minha família pela compreensão, confiança e pelo apoio em todos os momentos da minha vida, em especial este que é uma grande etapa da minha acadêmica.

Agradeço a Antonio F. L. Jacob Jr., meu orientador, pelas oportunidades dadas desde a Graduação, por me incentivar a fazer o Mestrado, por ser paciente, compreensível e ser uma referência na área acadêmica.

Agradeço ao SCRC (*Social CRM Research Center*), onde este trabalho foi iniciado, e aos professores envolvidos neste projeto. Obrigada por todo o conhecimento compartilhado.

Agradeço aos meus amigos, em especial as amizades construídas durante o Mestrado e que eu vou levar para a vida inteira. A amizade e o apoio de vocês foi muito importante nesta caminhada.

RESUMO

A Gestão de Relacionamento com o Cliente (CRM) tem sido um importante aliado para empresas dispostas a melhorar suas experiências com o cliente. Ultimamente, o Aprendizado de Máquina (AM) e seus algoritmos vêm transformando a forma como as empresas interagem com seus clientes por meio da análise de dados. Esta dissertação revisa a literatura sobre a aplicação de algoritmos de AM para melhorar processos de CRM e fornece uma visão geral de algoritmos usadas e sua aplicação a cada dimensão e elemento de CRM. Além disso, as implicações práticas são analisadas com base nos recentes desenvolvimentos de ferramentas de CRM no campo de AM. Esta dissertação fornece *insights* para estudos futuros, bem como para ferramentas de CRM que desejam melhorar seus serviços analíticos e automatizados.

Palavras Chave: Aprendizado de Máquina, CRM, Sistemas de Informação.

ABSTRACT

Current Customer Relationship Management (CRM) has been an important ally to companies willing to improve their customer experiences. Lately, Machine Learning (ML) and its algorithms have been transforming the way companies interact with their customers through data analysis. This dissertation reviews the literature on the application of ML algorithm to improve CRM processes and provides an overview of used techniques and their application to each CRM dimension and element. In addition, the practical implications are analyzed based on recent developments of CRM tools in the field of ML. This dissertation delivers insights for future studies as well as for CRM tools willing to improve their analytical and automated services.

Keywords: *Machine Learning, CRM, Information Systems.*

LISTA DE ILUSTRAÇÕES

Figura 1 – Tipos de CRM.	16
Figura 2 – Dimensões e Elementos do CRM Analítico.	17
Figura 3 – Infográfico de Inteligência Artificial.	20
Figura 4 – Ciclo do Processo KDD.	21
Figura 5 – Natureza Multidisciplinar da Mineração de Dados.	23
Figura 6 – Definição do Protocolo	29
Figura 7 – Processo de Seleção dos Estudos.	32
Figura 8 – Distribuição dos Estudos por Ano.	37
Figura 9 – Nacionalidade do Primeiro Autor.	38

LISTA DE TABELAS

Tabela 1 – <i>String</i> de Busca para a RSL	31
Tabela 2 – Resultado da Busca Automática.	35
Tabela 3 – Resultado Final da Seleção.	36
Tabela 4 – Total de Artigos Relevantes.	36
Tabela 5 – Classificação dos Algoritmos de AM em Relação aos Elementos de CRM. .	39
Tabela 6 – Número de Estudos Primários de Acordo com as Dimensões e Elementos do CRM.	43
Tabela 7 – Estudos Selecionados e seus Respectivos Algoritmos.	44
Tabela 8 – Algoritmos de AM Mais Usados nos Estudos Primários Selecionados. . . .	48
Tabela 9 – Áreas e Possíveis Casos de Uso de AM Valiosos para CRM.	49
Tabela 10 – Implicações Práticas da Literatura e Dimensões do CRM Correspondentes. .	50
Tabela 11 – Utilização de AM em Ferramentas de CRM.	53
Tabela 12 – Taxonomia de Algoritmos de Inteligência Computacional e de Aprendizado de Máquina Usadas Neste Trabalho.	59
Tabela 13 – Lista Final dos Estudos Primários	74

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
AD	Árvore de Decisão
AG	Algoritmo Genético
CRM	<i>Customer Relationship Management</i>
EQM	Erro Quadrático Médio
GB	<i>Gradient Boosting</i>
IA	Inteligência Artificial
IC	Inteligência Computacional
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
LR	<i>Logistic Regression</i>
KDD	<i>Knowledge Discovery in Databases</i>
K-NN	<i>K-Nearest Neighbor</i>
NB	<i>Naïve Bayes</i>
MD	Mineração de Dados
MK-STM	<i>Multi-kernel Support Tensor Machine</i>
MK-SVR	<i>Multi-kernel Support Vector</i>
PCA	<i>Principal Components Analysis</i>
RF	<i>Random Forest</i>
RM	<i>Relationship Management</i>
RNA	Redes Neurais Artificiais
RSL	Revisão Sistemática da Literatura
SOM	<i>Self-organizing Maps</i>
SVM	<i>Support Vector Machine</i>
TIC	Tecnologia da Informação e Comunicação
XGBoost	<i>Gradient Boosting</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	12
1.1.1	Objetivo Geral	12
1.1.2	Objetivos Específicos	13
1.2	Contribuição do Trabalho	13
1.3	Estrutura do Trabalho	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	<i>Customer Relationship Management</i>	15
2.1.1	Tipos de CRM	16
2.2	Inteligência Artificial	19
2.3	Inteligência Computacional	20
2.4	Descoberta do Conhecimento	21
2.4.1	Pré-processamento	22
2.4.1.1	Seleção de Dados	22
2.4.1.2	Limpeza de Dados	22
2.4.1.3	Codificação dos Dados	22
2.4.1.4	Transformação e Enriquecimento dos Dados	22
2.4.2	Mineração de Dados	23
2.4.2.1	Tarefas de Mineração de Dados	24
2.4.3	Pós-processamento	25
2.5	Aprendizado de Máquina	25
2.5.1	Avaliação do Modelo	26
3	METODOLOGIA	28
3.1	Planejamento da Revisão Sistemática da Literatura	29
3.1.1	Questões de Pesquisa	29
3.1.2	Base de Dados e Estratégias de Busca	30
3.1.3	Estratégia de Seleção	31
3.1.3.1	Critérios de Inclusão	32
3.1.3.2	Critérios de Exclusão	32
3.1.4	Avaliação de Qualidade do Estudo	33
3.1.5	Estratégia de Extração	34
3.1.6	Síntese e Análise dos Dados	34
4	CONDUÇÃO DA RSL	35
4.1	Resultados da Seleção	35
4.2	Avaliação de Qualidade	36
4.3	Resultados da Extração	37
5	ANÁLISE E DISCUSSÃO DOS RESULTADOS	39

5.1	Resultados da Revisão Sistemática	39
5.1.1	<i>QP1</i> . Quais áreas do CRM atualmente tem mais aplicações com AM?	39
5.1.2	<i>QP2</i> . Quais algoritmos de Aprendizado de Máquina estão sendo aplicados em processos de CRM?	43
5.1.3	<i>QP3</i> . Como os algoritmos de Aprendizado de Máquina melhoram os processos de CRM?	48
6	CONCLUSÕES	56
6.1	Limitações e Trabalhos Futuros	60
	REFERÊNCIAS	61
A	EXTRAÇÃO DOS DADOS DOS ESTUDOS PRIMÁRIOS SELECIONADOS	73
B	LISTA DE ESTUDOS RETORNADOS NA ETAPA DE SELEÇÃO . . .	81
C	DETALHES DA AVALIAÇÃO DE QUALIDADE	87

1 INTRODUÇÃO

As empresas estão constantemente em busca de meios para diferenciar suas marcas, produtos e serviços. Criar uma experiência melhor e eficaz para o cliente ao longo de toda a cadeia de valor de uma forma rentável é um dos principais desafios para as empresas hoje em dia (LEMON; VERHOEF, 2016; DELOITTE, 2016).

Ao projetar estratégias baseadas em dados para criar uma experiência única do cliente, as empresas confiam nos conceitos do campo de *Customer Relationship Management* (CRM). O CRM "utiliza a tecnologia como um facilitador para capturar, analisar e disseminar dados atuais e futuros dos clientes para identificar as necessidades dos clientes com mais precisão e desenvolver relacionamentos perspicazes" (BULL, 2003). Isso implica no alinhamento e na integração de objetivos estratégicos com processos de negócios apropriados e Tecnologia da Informação e Comunicação (TIC) (PAYNE; FROW, 2005). A base do CRM é a necessidade de amplo conhecimento sobre os clientes e a capacidade de enriquecer e utilizar esse conhecimento em processos operacionais e analíticos.

Conceitos como Inteligência Artificial (IA), Inteligência Computacional (IC), Aprendizado de Máquina (AM) e Mineração de Dados (MD) trouxeram ao mercado algoritmos e técnicas para processar e utilizar dados. IA é uma ciência multidisciplinar com uma ampla variedade de sub-campos que procura desenvolver e aplicar técnicas computacionais que simulam o comportamento humano em atividades específicas (RUSSELL; NORVIG, 2016). IC, no entanto, tem como objetivo usar a computação de aprendizado, adaptativa ou evolucionária para criar softwares que, em certo sentido, são inteligentes (BEGG; PALANISWAMI, 2006).

Com a IA como aliada, é possível melhorar vários processos de CRM, como: estratégias de vendas, satisfação do cliente, comunicação empresa-cliente, atração de novos clientes, etc.

Estudos anunciam o AM como uma questão fundamental para negócios futuros (IDOINE et al., 2018; COURT, 2015). No entanto, enquanto alguns estudos investigam o papel da MD no CRM, a literatura carece de informações sobre o estado da arte da pesquisa em relação aos campos de aplicação e benefícios do AM para o CRM. Esta dissertação analisa a literatura atual para identificar as técnicas e aplicações de AM aplicadas a processos de CRM e discute as implicações práticas dessas aplicações em ferramentas de CRM. Destina-se a identificar lacunas de pesquisa e melhorar as ferramentas existentes com base no cenário atual.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo desta revisão sistemática é identificar lacunas de pesquisa em aplicações de AM aplicadas em processos de CRM. Para assim, discutir as implicações práticas das aplicações

de AM em ferramentas de CRM.

Este estudo teve como objetivo realizar uma revisão sistemática da literatura de aplicações de aprendizado de máquina em áreas do CRM.

1.1.2 Objetivos Específicos

1. Realizar uma análise quantitativa e qualitativa da aplicação de algoritmos de aprendizado de máquina nos processos de CRM.
2. Analisar quais algoritmos de AM estão sendo usados, no contexto do CRM.

Com a realização deste trabalho, busca-se oferecer aos pesquisadores e profissionais da área de CRM, benefícios para seus trabalhos futuros. Trata-se de um trabalho de pesquisa interdisciplinar com o objetivo de contribuir com a área de CRM por meio de uma visão ampla das aplicações de algoritmos de Inteligência Artificial e Aprendizado de Máquina.

1.2 Contribuição do Trabalho

A principal contribuição deste trabalho é de informar a atual situação da pesquisa na área de CRM com relação a aplicações que envolvem aprendizado de máquina.

Com uma abordagem sobre as áreas de CRM e os algoritmos de aprendizado de máquina, serão detalhados quais dimensões do CRM estão sendo mais exploradas, quais algoritmos estão sendo mais utilizados, quais são os software de CRM que já fornecem serviços que utilizam aprendizado de máquina.

O trabalho foi desenvolvido de forma a apoiar tanto a academia quanto o ambiente corporativo, apresentando informações sobre aplicações já existentes, comparativos sobre quais problemas são mais abordados e quais áreas necessitam serem abordadas.

1.3 Estrutura do Trabalho

Esta dissertação está estruturada da seguinte forma:

Capítulo 2 - Fundamentação Teórica: Apresenta os principais conceitos relacionados a CRM, inteligência artificial, inteligência computacional, descoberta do conhecimento e aprendizado de máquina;

Capítulo 3 - Metodologia: Apresenta o método da Revisão Sistemática da Literatura. Detalha o protocolo da RSL, descrevendo os procedimentos para a busca, seleção e análise dos estudos;

Capítulo 4 - Condução da Revisão Sistemática da Literatura: Expõe o resultado de cada etapa da RSL, assim são apresentados dados que vão servir de base para responder as perguntas de pesquisa.

Capítulo 5 - Análise e Discussão dos Resultados: Apresenta o resultado da RSL por completo e comenta as questões de pesquisa.

Capítulo 6 - Conclusões: Exibe as contribuições e limitações da pesquisa e recomendações para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta conceitos básicos envolvidos na realização deste trabalho, incluindo: CRM, inteligência artificial, inteligência computacional, mineração de dados e aprendizado de máquina.

2.1 *Customer Relationship Management*

Nos primeiros estágios do comércio, os clientes eram conhecidos, por isso as pessoas de vendas sabiam o que os clientes preferiam e quem pagava, e quem geralmente pagava a tempo. No entanto, quando as empresas cresceram, as informações sobre clientes individuais foram perdidas entre as massas. Para lidar com as bases de clientes sempre em expansão, grupos de clientes ou *marketing* segmentado foram criados, onde os clientes são colocados em *slots* que melhor se encaixam em seus desejos. Usando esses *slots* ou grupos de clientes, a maioria dos requisitos e expectativas poderia ser atendida, mas há comprometimento em muitos desejos, já que os serviços são adaptados a grupos, não a indivíduos (BOSE, 2002).

Para lidar com esse problema, o termo "*Relationship Management*" começou a surgir por volta de 1980. A ideia era trabalhar mais o relacionamentos diretos com os clientes. As empresas perceberam que isso permitia que elas aprendessem mais e se adaptassem melhor a seus clientes, além de criar valor adicional para seus clientes e para eles mesmos (BERFENFELDT, 2010). Contudo, a empolgação com o *marketing* de relacionamento logo desapareceu mais uma vez, já que estava se mostrando difícil conseguir bons resultados a curto prazo com ele. E as estratégias e sistemas usados para tentar manter esses bancos de dados de clientes também se mostraram muito caros e tediosos para manter (PEELEN, 2005; XU et al., 2002). Em várias empresas isso pode ter sido o fim do Gerenciamento de Relacionamento, mas com a introdução de Tecnologias de Informação mais avançadas, ele voltou com um novo nome, *Customer Relationship Management* (PEELEN, 2005).

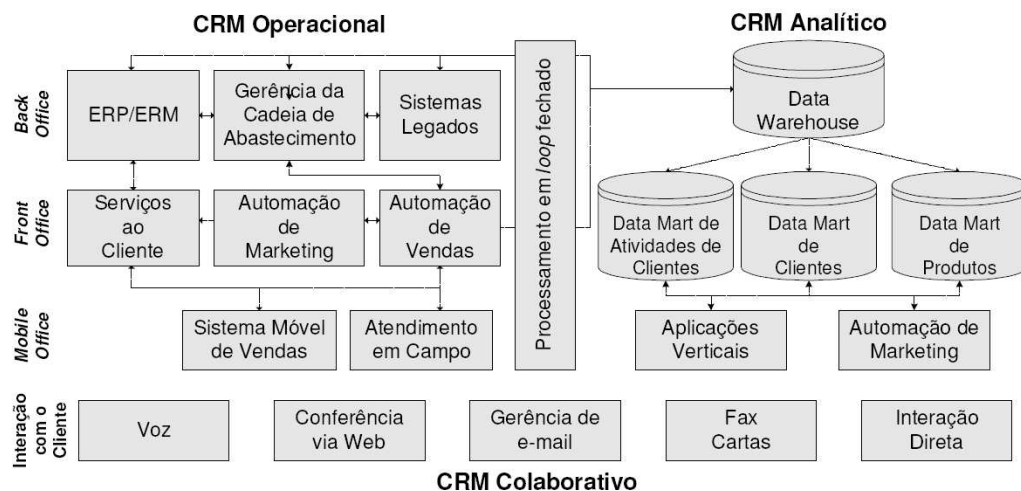
Payne e Frow (2005) resumiram a literatura de CRM com o propósito de ter uma definição mais concisa, e assim o definiram como:

O CRM é uma abordagem estratégica que se preocupa em criar valor para o acionista aprimorado por meio do desenvolvimento de relacionamentos apropriados com os principais clientes e segmentos de clientes. O CRM une o potencial das estratégias de *marketing* de relacionamento e TI para criar relacionamentos rentáveis e de longo prazo com os clientes e outras partes interessadas. O CRM oferece melhores oportunidades para usar dados e informações para entender os clientes e criar valor com eles. Isso requer uma integração interfuncional de processos, pessoas, operações e recursos de *marketing* que é ativada por meio de informações, tecnologia, e aplicações (PAYNE; FROW, 2005).

2.1.1 Tipos de CRM

Assim como o conceito, diferentes autores defendem diferentes classificações para o CRM. Para Payne e Frow (2005), Peppers e Rogers (2000), Khodakarami e Chan (2014), Greenberg (2001) existem três tipos: operacional, colaborativo e analítico. A Figura 1 ilustra a composição dos tipos.

Figura 1 – Tipos de CRM.



Fonte: (PEPPERS; ROGERS, 2000).

CRM Operacional

Conforme Peppers e Rogers (2000), o CRM operacional refere-se ao componente que ajuda a melhorar a eficiência das operações diárias dos clientes. É a área que está relacionada com a automatização de processos de negócios que envolvem pontos de contato com o cliente. Incluindo automatização de vendas, de *marketing* e atendimento ao cliente. Portanto é para suprir o setor de atendimento como, por exemplo, os *Call Centers* (PAYNE; FROW, 2005).

Para garantir o sucesso do CRM operacional, as empresas devem se concentrar nos requisitos dos clientes e os funcionários devem ter as habilidades certas para satisfazê-los. A saída das soluções operacionais de CRM geralmente é apenas um nível de resumo, mostrando quais atividades ocorreram, mas não explicando suas causas ou impacto (REYNOLDS, 2002).

CRM Analítico

De acordo com Greenberg (2001), o CRM analítico são procedimentos que envolvem a captação, o armazenamento, o acesso, o processamento e a transmissão dos dados, nas interações entre a empresa e seus clientes.

E para (SWIFT, 2001 apud OLIVEIRA et al., 2012), o CRM analítico é categorizado em quatro dimensões: (i) Identificação do Cliente (*Customer Identification*), (ii) Atração do Cliente (*Customer Attraction*), (iii) Retenção do Cliente (*Customer Retention*) e (iv) Desenvolvimento do Cliente (*Customer Development*). Essas quatro dimensões podem ser vistas como

um ciclo fechado do sistema de gestão de clientes (AU; CHAN; YAO, 2003; KRACKLAUER; MILLS; SEIFERT, 2004; LING; YEN, 2001). A Figura 2 ilustra as dimensões e seus respectivos elementos.

Figura 2 – Dimensões e Elementos do CRM Analítico.



Fonte: Elaborado pela Autora.

As quatro dimensões são descritas a seguir:

- i. Identificação do cliente: a identificação do cliente inclui principalmente segmentação de clientes (*customer segmentation*) e análise de clientes-alvo (*target customer analysis*). A segmentação de clientes implica a subdivisão do conjunto de todos os clientes em segmentos menores, incluindo clientes com características semelhantes. A análise de clientes-alvo envolve a definição dos segmentos mais atraentes para a empresa, com base nas características dos clientes. A seleção dos grupos-alvo requer a coleta de dados

quantitativos e qualitativos sobre esses grupos (WOO; BAE; PARK, 2005 apud OLIVEIRA et al., 2012).

- ii. Atração do cliente: esta etapa segue a identificação dos clientes. Tendo identificado os grupos-alvo, as empresas concentram esforços e alocam recursos para atrair esses segmentos. Vantagens competitivas, como preço e outras características de diferenciação, podem ser impulsionadores da atração dos clientes. Outro driver de atração do cliente é o *marketing* direto (*direct marketing*). Esse é um elemento do mix de *marketing* da empresa que motiva os clientes a fazer um pedido imediatamente. Por exemplo, mala direta ou distribuição de cupons são exemplos típicos de *marketing* direto. A atração do cliente envolve o uso de um método apropriado de comunicação e a eliminação de qualquer tipo de esforço desperdiçado (CHEUNG et al., 2003; LIAO; CHEN, 2004; HE et al., 2005; PRINZIE; POEL, 2005; KRACKLAUER; MILLS; SEIFERT, 2004; OLIVEIRA et al., 2012).
- iii. Retenção do cliente: essa dimensão é uma das principais preocupações do CRM. a satisfação do cliente é a principal questão relacionada à retenção de clientes. A satisfação do cliente pode ser definida como a comparação das expectativas dos clientes (resultantes do padrão pessoal, imagem da empresa, conhecimento de alternativas, etc.) com as percepções (resultantes da experiência real, impressão subjetiva do desempenho do produto, adequação do produto ou serviço, etc). A percepção do cliente sobre o valor oferecido pela empresa leva à retenção sustentada de clientes. Além disso, uma experiência de compra de alta qualidade leva a um sentimento emocional positivo, que permite à empresa alcançar a lealdade do cliente desejada. Os elementos dessa dimensão de CRM incluem: *marketing* personalizado (*marketing one-to-one*), programas de fidelidade e bônus (*loyalt program*) e gerenciamento de reclamações (*complaint management*). O *marketing* personalizado envolve campanhas de marketing personalizadas com suporte na análise, detecção e previsão de mudanças no comportamento do cliente. Os programas de fidelidade e bônus envolvem campanhas ou atividades de apoio que visam manter um relacionamento de longo prazo com os clientes. Exemplos de programas de fidelidade incluem pontuação de crédito, qualidade de serviço ou satisfação e análise de churn, ou seja, análise se um cliente provavelmente sairá para um concorrente (KRACKLAUER; MILLS; SEIFERT, 2004; JIANG; TUZHILIN, 2006; NGAI; XIU; CHAU, 2009; OLIVEIRA et al., 2012).
- iv. Desenvolvimento do cliente: o foco principal dessa dimensão é aumentar intensidade da transação, valor da transação e lucratividade individual do cliente. Os principais elementos do desenvolvimento do cliente são a análise do valor da vida útil do cliente (*customer lifetime cycle*) e o *up/cross-selling*. O valor da vida útil do cliente é a receita líquida total que uma empresa pode esperar de um cliente. *Up/cross selling* são as atividades promocionais que visam aumentar o número de serviços ou produtos associados ou intimamente relacionados que um cliente usa em uma empresa. O *design* de tais atividades

promocionais é geralmente suportado pela análise de afinidade (*market basket analysis*), que permite identificar os padrões subjacentes ao comportamento do cliente (DREW et al., 2001; AGGARWAL; PROCOPIUC; YU, 2002; GIRAUD-CARRIER; POVEL, 2003; KUBAT et al., 2003; OLIVEIRA et al., 2012).

CRM Colaborativo

O CRM colaborativo pode ser visto como um centro de comunicação que fornece a conexão entre as empresas e seus clientes, fornecedores e parceiros de negócios. De fato, permite que clientes, funcionários, vendedores e parceiros acessem, distribuam e compartilhem dados. Publicações personalizadas, e-mail, comunidades, conferências e centros de interação com relacionamento via web são exemplos de serviços colaborativos. Esses serviços tornam o trabalho em equipe mais fácil e produtivo, permitindo que as empresas aprimorem os processos e, conseqüentemente, melhorem a satisfação dos clientes. Sendo assim, o CRM de desenvolvimento é usado para estabelecer o valor vitalício dos clientes além da transação, criando um relacionamento de parceria (GREENBERG, 2001 apud OLIVEIRA et al., 2012).

2.2 Inteligência Artificial

A Inteligência Artificial tenta entender e construir entidades inteligentes que pensam e agem racionalmente como seres humanos para resolver problemas ou tomar decisões (STUART; PETER et al., 2003). De acordo com Zuylen (2012), IA também significa sistemas de computador que demonstram sistema de vida complexos, como comportamentos. Por exemplo, eles poderiam imitar o enxame de insetos, a colônia de formigas, a microbiologia ou um sistema neural.

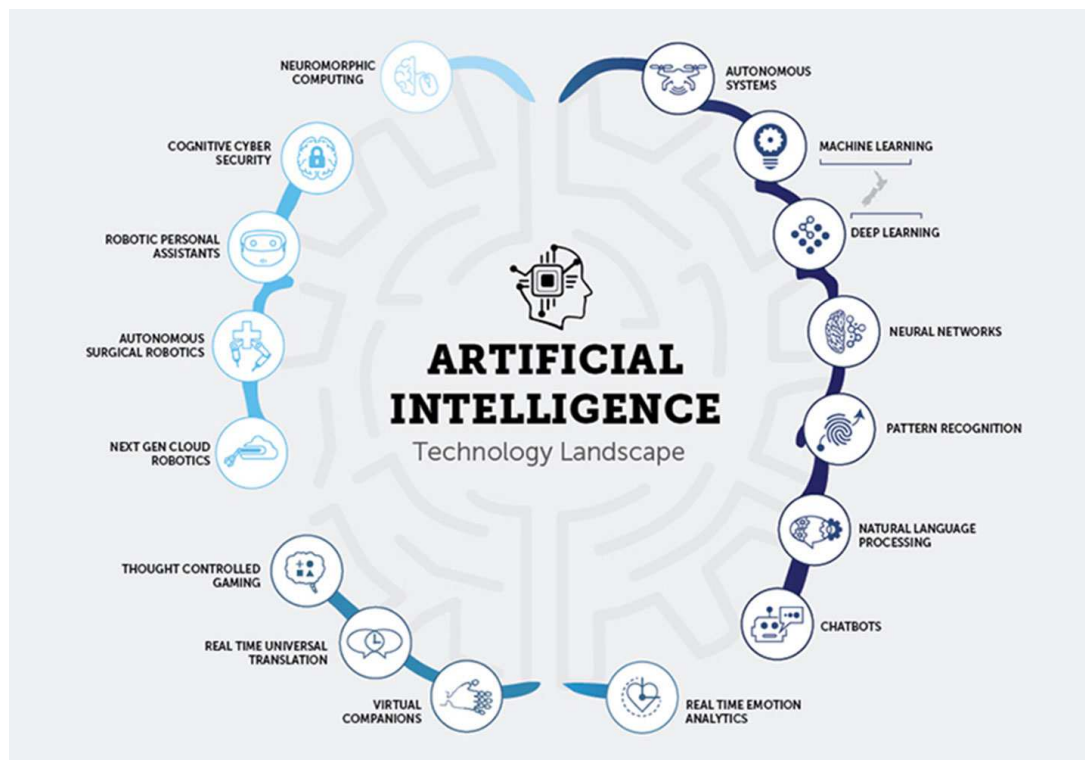
As aplicações de IA abrangem vários domínios, incluindo controle autônomo, robótica, compreensão de linguagem e visão computacional. As aplicações podem ser encontradas em vários campos, como economia, manufatura, engenharia e medicina (ZUYLEN, 2012). A Figura 3 ilustra os domínios de conhecimento que compõem a inteligência artificial.

Segundo Sadek (2007), os métodos de IA incluem duas categorias principais - IA simbólica e inteligência computacional. IA simbólica se concentra no desenvolvimento de sistemas baseados em conhecimento capazes de tomar decisões em um domínio particular, utilizando o conhecimento de um especialista humano. Inteligência Computacional (IC) consiste em métodos como sistema fuzzy, rede neural e computação evolutiva. A diferença entre a IC e a IA simbólica é que a saída é gerada sem o uso da base de conhecimento, como regras, quadros ou casos.

Os fatores mais influentes no surgimento da IA agora são capacidade e potência de computação, dados e algoritmos. O poder e a capacidade dos processadores aumentaram significativamente. Atualmente, há uma enorme quantidade de dados disponíveis sobre o clima, mídias

¹ Disponível em: <https://www.callaghaninnovation.govt.nz/blog/ai-demystified>

Figura 3 – Infográfico de Inteligência Artificial.

Fonte: Site Callaghan Innovation¹

sociais e ciências médicas, e as máquinas finalmente podem explorar esses dados. Ao mesmo tempo, os custos de armazenamento para gerenciamento de dados caíram e o desenvolvimento com armazenamento de dados levou a uma maneira mais rápida de analisar grandes quantidades de dados (POTTALA et al., 2018 apud VÄHÄKAINU; NEITTAANMÄKI, 2018).

2.3 Inteligência Computacional

Apesar de ser muito difundida, a Inteligência Computacional não tem uma definição comumente aceita. Entretanto, existem algumas definições. Conforme o IEEE (2008), a IC é um ramo do estudo da IA. A comunidade internacional de líderes profissionais e pesquisadores em inteligência computacional “IEEE Computational Intelligence Society” sinala que a inteligência computacional visa o uso de aprendizado, adaptativo ou de computação evolutiva para criar *software* que é, em algum sentido, inteligente.

Outra definição é dada por Pérez (2010), onde afirma que a IC é um sistema de métodos e infraestruturas que melhoram a inteligência humana, aprendendo e descobrindo novos padrões em ambientes dinâmicos para resolver problemas práticos, sendo uma ferramenta colaborativa com o ser humano, conduzindo a melhorar a inteligência humana.

A partir das definições de IC, Rezende (2003), declara que as pesquisas em inteligência computacional tem por objetivo, capacitar o computador a executar funções que são empenhadas

pelo ser humano usando o conhecimento e o raciocínio.

A IC considera uma coleção de paradigmas computacionais com inspiração biológica e linguística, em que a teoria, o modelo, a implementação e o desenvolvimento de redes neurais (incluindo também máquinas de vetores de suporte, por semelhança), algoritmos evolutivos, sistemas baseados em lógica *fuzzy* e sistemas inteligentes híbridos (MAITA, 2016 apud ONG; LIM; CHEN, 2010).

2.4 Descoberta do Conhecimento

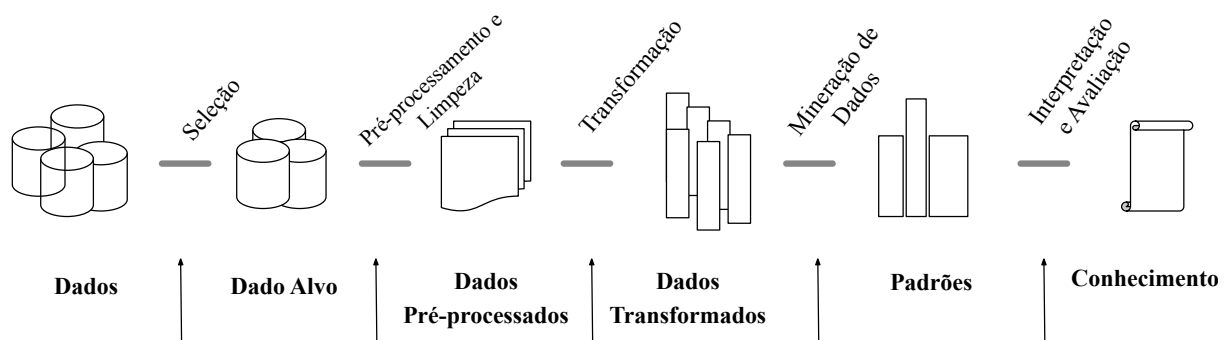
A informação é o ativo mais importante para os negócios das organizações, tornando-se algo essencial para ganho de competitividade entre as empresas de pequeno, médio e grande porte. As estratégias assumidas para tal ganho devem basear-se em informações concretas, visando uma minimização na ocorrência de erros para a tomada de decisões por parte dos gestores (DANTAS et al., 2008).

O grande problema se apresenta na dificuldade de converter os dados, que essas organizações obtêm, em informação útil. Em meio deste problema surgiu um processo chamado de Descoberta do Conhecimento em Banco de Dados (*Knowledge Discovery in Databases - KDD*).

Processos KDD visam disponibilizar e integrar todos os recursos necessários para as atividades de transformação de dados brutos em conhecimento útil, e funcionalidades para localização, carga, preparação dos dados e ferramentas para a execução de análises (VALENTIN, 2006).

De acordo com Fayyad et al. (1996), as etapas do processo são: seleção, pré-processamento e limpeza, transformação, mineração de dados (data mining) e interpretação/avaliação. Como se pode ver, o processo compreende, na verdade, todo o ciclo que o dado percorre até resultar em conhecimento (ver Figura 4).

Figura 4 – Ciclo do Processo KDD.



Fonte: Adaptado de (PEPPERS; ROGERS, 2000).

Fayyad et al. (1996), define o processo como não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados

em um banco de dados; iterativo, sendo assim o usuário pode interferir no curso das atividades; iterativo, onde o resultado de cada uma é dependente dos resultados das que a precedem.

Os autores Goldschmidt e Passos (2005) agrupam ainda mais as etapas tornando o processo KDD dividido em três etapas, sendo estas: pré-processamento, mineração de dados e pós-processamento. A seguir, as etapas de pré-processamento e pós-processamento serão detalhadas e a etapa de mineração de dados será abordada na seção seguinte.

2.4.1 Pré-processamento

Esta etapa inclui operações de manipulação de dados, que envolve tarefas como remover ruídos ou subcamadas, coletando informações necessárias para tarefas de descoberta do conhecimento. Nesta etapa, também são decididas estratégias para tratar campos que não são necessários no processo de mineração de dados.

Goldschmidt e Passos (2005) descrevem os itens que compõem esta operação:

2.4.1.1 Seleção de Dados

Compreende a identificação de quais informações devem ser consideradas durante o processo de KDD, sendo um processo bastante complexo, uma vez que os dados podem vir de uma série de fontes diferentes (data warehouses, planilhas, sistemas legados) e podem possuir os mais diversos formatos. Este passo possui impacto significativo sobre a qualidade do resultado do processo.

2.4.1.2 Limpeza de Dados

Esta é uma parte crucial no processo, pois a qualidade dos dados vai determinar a eficiência dos algoritmos de mineração. Nesta etapa deverão ser realizadas tarefas que eliminem dados redundantes e inconsistentes, recuperem dados incompletos e avaliem possíveis dados discrepantes ao conjunto (outliers). Envolve qualquer tipo de tratamento referente aos dados selecionados de forma a assegurar a qualidade da informação por eles representada.

2.4.1.3 Codificação dos Dados

os dados devem ser codificados de forma que possam ser utilizados como entrada para os algoritmos de mineração de dados. Por exemplo, transformar valores numéricos em categorias ou intervalos; ou o inverso, como representar numericamente valores de atributos categóricos.

2.4.1.4 Transformação e Enriquecimento dos Dados

consiste em reduzir as variáveis envolvidas por atributos que representem as características principais dos dados das bases. Já o enriquecimento é um item onde se buscam mais informações que possam ser agregadas a base de dados, para prover melhores resultados. Para

isso, podem-se realizar consultas a bases externas de dados, pesquisas de campo e outras técnicas (GOLDSCHMIDT; PASSOS, 2005).

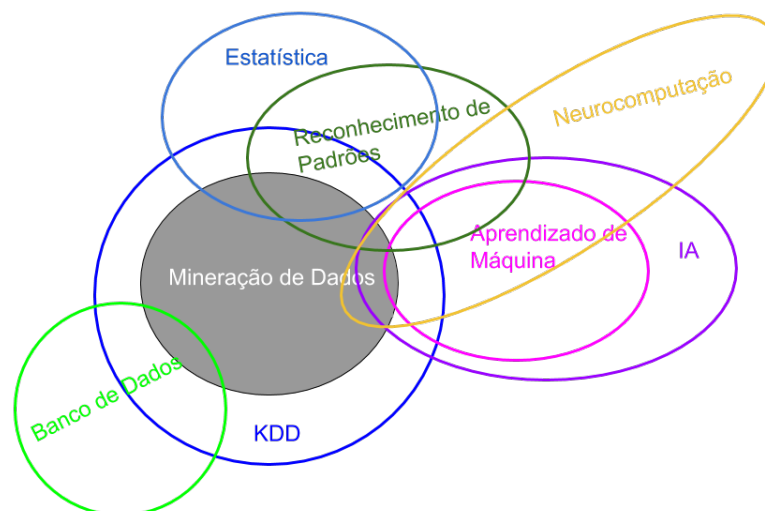
2.4.2 Mineração de Dados

Os termos mineração de dados (MD) e KDD são muitas vezes confundidos. O KDD foi proposto em 1995 para descrever todo o processo de extração de conhecimento de dados. Neste contexto, conhecimento significa relacionamentos e padrões entre elementos de dados. MD deve ser usado exclusivamente para a fase de descoberta do processo KDD (SUMATHI; SIVANANDAM, 2006).

A área de pesquisa que relaciona MD e descoberta do conhecimento é resultado do esforço para satisfazer a demanda de transformar dados em conhecimento útil e orientado a tarefas. Com isso, os pesquisadores têm explorado ideias e métodos desenvolvidos em aprendizagem de máquina, reconhecimento de padrões, análise de dados estatísticos, visualização de dados, redes neurais, etc. (SUMATHI; SIVANANDAM, 2006).

MD é uma disciplina composta que se sobrepõe a outros ramos da ciência. A Figura 5 mostra as contribuições de muitos campos diferentes no desenvolvimento da ciência da mineração de dados (DEAN, 2014).

Figura 5 – Natureza Multidisciplinar da Mineração de Dados.



Fonte: Adaptado de (HALL et al., 2014).

Sumathi e Sivanandam (2006), afirmam que o objetivo da mineração de dados é extrair padrões implícitos, previamente desconhecidos e potencialmente úteis dos dados. A MD consiste em muitas técnicas como: classificação (*árvores de decisão, naïve bayes, k-nearest neighborhood, redes neurais*), agrupamento (*k-means, hierarchical clustering, density-based clustering*), associação (associação unidimensional, multidimensional, multinível, associação baseada em restrições).

Existem várias aplicações para minerações de dados , em diversas áreas. Alguns exemplos são: (i) agrupamento de clientes por perfis para gestão do relacionamento com os clientes; (ii) predição de séries temporais hidrológicas a partir de redes neurais (TAMPELINI et al., 2009); e (iii) classificação (MADEO et al., 2010).

De acordo com (HAN; KAMBER; PEI, 2006) a MD pode ser classificada em duas categorias: padrões descritivos e padrões preditivos. As tarefas descritivas caracterizam as propriedades gerais dos dados em uma base de dados. Já os padrões preditivos buscam através de inferência prever valores desconhecidos. Padrões preditivos são construídos para resolver problemas específicos de predição, em um ou mais atributos de um banco de dados. Ainda afirmam que esses padrões não precisam necessariamente prever o futuro, mas sim fazer a descoberta de um atributo desconhecido a partir de outros atributos conhecidos.

Em contrapartida, para John (1997) os padrões descritivos não têm por objetivo resolver um problema específico, mas representar padrões interessantes que o especialista em um domínio desconhece.

2.4.2.1 Tarefas de Mineração de Dados

As tarefas de mineração podem extrair diferentes tipos de conhecimento, sendo necessário decidir, no início do processo, qual o tipo de conhecimento que o algoritmo deve extrair (CASTANHEIRA, 2008).

Algumas subdivisões são classificadas nas categorias. Sendo a tarefa preditiva subdivida em classificação e regressão. A tarefa descritiva em regras de associação, agrupamento e sumarização (MARTINS, 2010).

MÉTODOS PREDITIVOS

- **Classificação:** função que mapeia (classifica) um registro de uma base de dados a uma ou mais classes (rótulos) predefinidos. Feito isso, é possível aplicar essa função a novos registros que venham ser incorporados à base de dados. As formas mais comuns de representação de conhecimento dos algoritmos de classificação são regras e árvores.
- **Regressão:** consiste na busca por uma função que mapeia os registros de um banco de dados em valores reais (GOLDSCHMIDT; PASSOS, 2005). Essas funções podem ser lineares ou não, e se restringem a dados numéricos.

MÉTODOS DESCRITIVOS

- **Regras de Associação:** tem por objetivo encontrar padrões frequentes entre conjunto de itens de dados e é muito usada no âmbito das aplicações comerciais na área de marketing.

- **Agrupamento:** é a classificação de objetos ou registros em diferentes grupos, ou seja, particionando do conjunto de dados em subconjuntos ou agrupamentos, assim cada subconjunto compartilha similaridades.
- **Sumarização:** envolve métodos para encontrar uma descrição das características comuns para conjuntos de dados ou agrupamentos obtidos através da técnica de agrupamento.

2.4.3 Pós-processamento

Esta etapa engloba a visualização, a análise e a interpretação das saídas geradas na etapa de mineração de dados. Com isso, é gerado um modelo de conhecimento, ou seja, qualquer abstração de conhecimento, expresso em alguma linguagem que descreva algum conjunto de dados (GOLDSCHMIDT; PASSOS, 2005)

2.5 Aprendizado de Máquina

O AM é uma metodologia que dá aos computadores a capacidade de aprender sem programação explícita. AM está incluído nos campos de IA ou IC, dependendo dos algoritmos ou técnicas utilizadas. Ele se baseia no princípio de que as máquinas devem aprender com os dados fornecidos, a partir de suas experiências. O AM usa diferentes técnicas do campo da MD, incluindo alguns algoritmos extras de aprendizado (KOHONEN, 1990; LATHAUWER; MOOR; VANDEWALLE, 1994; SAMUEL, 1959; BUCZAK; GUVEN, 2015; MISHRA; RANI, 2017).

Outra definição é apresentada por (FACELI et al., 2011), que define AM como um processo de indução de uma hipótese (ou aproximação de uma função) a partir da experiência passada.

Haykin et al. (2009) classifica as técnicas de aprendizado de máquina em dois tipos de estratégia de aprendizado: aprendizado supervisionado e aprendizado não supervisionado. Chapelle, Scholkopf e Zien (2006), concordam porém incluem o aprendizado semissupervisionado.

- **Aprendizado Supervisionado:** neste tipo de aprendizagem, o sistema recebe um conjunto de dados com diferentes valores de parâmetros de exemplo e decisões/classificação, a partir dos quais infere uma função matemática, que mapeia automaticamente um sinal de entrada para um sinal de saída. Desse modo, descobre o que é deve ser feito (THULIN, 2007). Assim, requer algum tipo de supervisão no processo onde possa ser informado os exemplos típicos de cada classe, os quais, dentro de um processo de treinamento, é permitido identificar a classe de novos objetos (Haykin, 1999). De acordo com os valores atribuídos à classe, o problema é conhecido como classificação ou regressão (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Árvores de decisão, *k-Nearest Neighbour* (*k-NN*), *Naïve Bayes* e *Support Vector Machine* (*SVM*) são exemplos de algoritmos que tratam este tipo de aprendizado de máquina.

- **Aprendizado Não Supervisionado:** o sistema atua e observa as consequências de suas ações, sem se referir a casos de tipo pré-definidos além daqueles previamente observados, ou seja, o sistema aprende com base em tentativa e erro (THULIN, 2007). O objetivo comum desse aprendizado é agrupar documentos semelhantes no mesmo *cluster*. Uma das possíveis vantagens da aplicação de métodos de agrupamento é que os dados de treinamento poderiam ser selecionados de forma mais adequada. De fato, em vez de tomar uma amostra aleatória do conjunto de resumos, pode-se considerar a realização de uma análise de *cluster* inicial e selecionar resumos de treinamento de cada um dos clusters resultantes (JASPERS; TROYER; AERTS, 2018).
- **Aprendizado Semissupervisionado:** O aprendizado semissupervisionado é a combinação entre o aprendizado não supervisionado e supervisionado. Na verdade, a maioria das estratégias de aprendizado semissupervisionados se baseia em ampliar o aprendizado não supervisionado ou supervisionado para incluir informações adicionais típicas do outro paradigma de aprendizado. Esse tipo de aprendizado tem conquistado espaço, porque promete maior precisão com menos esforço. Já que, em muitas tarefas, há escassez de dados rotulados. Os rótulos podem ser difíceis de serem obtidos seja porque exigem dispositivos especiais, experimentos caros e lentos ou humanos para rotular os dados. Desse modo, o aprendizado semissupervisionado pode atingir o mesmo nível de desempenho que o aprendizado supervisionado, mas com menos ocorrências rotuladas. (GOLDBERG; ZHU, 2010; ZHU; LAFFERTY; ROSENFELD, 2005).

2.5.1 Avaliação do Modelo

Após a construção do modelo é necessário avaliar o desempenho do classificador. Dependendo do contexto, o processo pode ser avaliado diferente. Um modelo produtivo é avaliado por meio de um conjunto de testes ou uma técnica de validação cruzada. Um conjunto de testes contém exemplos pré-classificados diferentes daqueles do conjunto de treinamento e é usado apenas para avaliação, não para treinamento. Se os dados são escassos, é sensato usar validação cruzada para não desperdiçar nenhum dado, o que poderia ser útil para melhorar o desempenho do classificador (SANTOS, 2018; THULIN, 2007).

Em problemas de regressão, a performance é avaliada por meio do erro quadrático médio (EQM), e é dado por:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Sua interpretação refere-se à distância média entre valores observados e os preditos pelo modelo. O EQM será pequeno se as respostas preditas pelo modelo forem muito próximas das observadas e será grande se, em algumas observações, a resposta predita e a observada diferirem substancialmente. Como o objetivo final da modelagem preditiva é obter predições acuradas

em novos dados, não utilizados para o ajuste do modelo, a performance preditiva do modelo selecionado deve ser avaliada a partir dos dados de teste (JAMES et al., 2013; SANTOS, 2018).

Para problemas de classificação as métricas comumente usadas, são: acurácia, precisão e *recall* (THULIN, 2007).

$$\text{Acuracia} = \frac{\text{Nº de amostras classificadas corretamente}}{\text{Nº total de amostras}}$$

$$\text{Precisao}(C) = \frac{\text{Nº de amostras classificadas corretamente como Classe } C}{\text{Nº total de amostras classificadas como Classe } C}$$

$$\text{Recall}(C) = \frac{\text{Nº de amostras classificadas como Classe } C}{\text{Nº total de amostras na Classe } C}$$

Acurácia é a media geral de desempenho do classificador; enquanto que a precisão lida com os erros do classificador, resultando em apenas dados classificados corretamente; e o *recall* é a frequência que o classificador identifica corretamente os exemplos de uma classe (THULIN, 2007).

3 METODOLOGIA

Este capítulo pretende descrever detalhadamente o método de pesquisa usado neste trabalho, a fim de aumentar a confiabilidade dos resultados alcançados, permitir a replicabilidade deste estudo e alcançar os objetivos.

Para Fonseca (2016) a metodologia é o estudo da organização, dos caminhos a serem percorridos, para se realizar uma pesquisa ou um estudo, ou para se fazer ciência.

Desse modo, este trabalho visa estudar de forma ampla o contexto atual de aplicações de algoritmos de AM nas áreas de CRM, desta forma é necessário o uso de métodos de pesquisa sistemáticos. De forma ampla, os métodos de pesquisa usados enquadra-se como:

- **Natureza básica:** uma vez que objetiva gerar conhecimento novo, útil para a ciência, neste caso, na área de CRM e Aprendizado de Máquina.
- **Gênero empírico:** já que o trabalho foi direcionado principalmente pela busca de dados relevantes obtidos por meio da experiência, da vivência de outros pesquisadores, publicadas por meio de trabalhos de pesquisa existentes na literatura científica; tendo como objetivo chegar em novas conclusões a partir da maturidade experimental desses trabalhos avaliados.
- **Propósito exploratório e descritivo:** primeiramente com caráter “exploratório” por proporcionar maior familiaridade com o problema, que pode ainda ser considerado novo, principalmente considerando o uso de técnicas de inteligência computacional e de aprendizado de máquina em aplicações de CRM, com vistas a torná-lo mais explícito ou construir hipóteses; mas também com caráter “descritivo” por já possuir o objetivo de descrever os fatos e/ou fenômenos da realidade estudada.
- **Abordagem mista qualitativa e quantitativa:** em relação a análise e apresentação dos resultados, sendo que, em alguns casos, os resultados podem ser mensurados em números, e então classificados e analisados com estatística descritiva básica, e, em outros, uma opinião crítica das lições aprendidas no processo é sustentada nas teorias encontradas.

A realização deste trabalho está fundamentada no estudo dos principais trabalhos já publicados na área de AM com aplicação em CRM; tendo como fonte de informação artigos já publicados nos principais veículos científicos internacionais. Sendo assim, esta dissertação tem como procedimento técnico uma pesquisa, principalmente, do tipo "bibliográfica". Para a elaboração deste, foi utilizado o método de pesquisa "revisão sistemática da literatura", baseada nas diretrizes estabelecidas por (KITCHENHAM; CHARTERS, 2007).

A revisão sistemática é um meio de identificar, avaliar e interpretar toda a pesquisa disponível relevante para uma pesquisa específica, ou área temática ou fenômeno de interesse (KITCHENHAM, 2004). A autora ainda complementa, que os estudos individuais que contri-

buem para uma revisão sistemática são chamados de estudos primários; uma revisão sistemática é uma forma de estudo secundário.

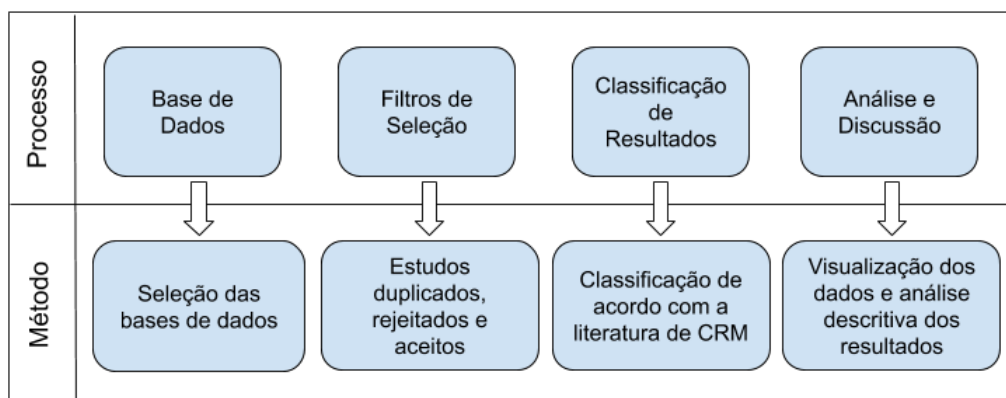
Randolph (2009) apresenta algumas razões para uma revisão sistemática ser realizada: delimitar o problema de pesquisa, buscar novas linhas de investigação, evitar abordagens infrutíferas, buscar conhecimentos metodológicos, identificar recomendações para futuras pesquisa e buscar fundamentação teórica.

Segundo Kitchenham (2004), a revisão sistemática envolve várias atividades discretas, que podem ser agrupadas em três fases principais: (i) planejamento, deve assegurar que a revisão sistemática é realmente necessária, e um protocolo de revisão deve ser desenvolvido; (ii) condução, que inclui a identificação da pesquisa, seleção dos estudos, avaliação da qualidade dos estudos, extração de dados, e a organização da informação extraída; e (iii) relatório, que inclui a documentação dos resultados em artigos científicos e trabalhos de dissertação ou teses.

3.1 Planejamento da Revisão Sistemática da Literatura

As próximas seções descrevem detalhadamente o protocolo da RSL. Sendo assim, as questões de pesquisa, base de dados e estratégias de busca, estratégia de seleção de trabalhos, avaliação de qualidade, estratégias de extração e síntese e análise de dados dos estudos são definidos. A Figura 6 apresenta todas as etapas para atingir os resultados desejados nesta RSL e, logo mais, estas etapas são descritas.

Figura 6 – Definição do Protocolo



Fonte: Elaborado pela Autora.

Uma descrição resumida do protocolo definido e usado neste trabalho para a realização da revisão sistemática da literatura, encontra-se em um artigo já publicado em uma conferência internacional (CHAGAS et al., 2018).

3.1.1 Questões de Pesquisa

QP1. Quais áreas do CRM atualmente tem mais aplicações com AM?

Com o intuito de direcionar essa questão, foram consideradas as áreas de CRM de acordo com (SWIFT, 2001), descrito no capítulo 2, seção 2.1.

QP2. Quais algoritmos de Aprendizado de Máquina são aplicados em processos de CRM?

Para responder essa pergunta, os conceitos de AM e CRM foram definidos, juntamente com as suas dimensões e elementos. Com isso, partiu-se para a revisão sistemática dos artigos que apresentavam alguma relação prática de algoritmos de AM em CRM.

QP3. Como os algoritmos de Aprendizado de Máquina melhoram os processos de CRM?

Para responder essa questão, as principais ferramentas de CRM foram analisadas a fim de obter uma visão geral, observando quais já investem em AM e em quais processos os algoritmos são aplicados.

3.1.2 Base de Dados e Estratégias de Busca

Pesquisas sobre os tópicos de AM e CRM podem estar espalhadas em diferentes revistas de diferentes áreas, como: Tecnologia da Informação, Sistemas de Informação, *Marketing*, Administração, etc. Portanto, a fim de maximizar a quantidade de trabalhos encontrados e apresentar resultados de diferentes tipos de bancos de dados, a pesquisa foi conduzida através de bases de dados eletrônicas, já que são mais acessíveis e permitem fácil replicabilidade do estudo. Assim foram definidas quatro bases:

- Ebsco Database
- IEEE Xplore
- Science Direct
- Emerald

As bases de dados selecionadas forneceram uma bibliografia abrangente da literatura acadêmica em todas as disciplinas relevantes dos tópicos de AM e CRM. A biblioteca digital Science Direct indexou mais de 60% artigos.

A fim de encontrar trabalhos relacionados a CRM que utilizam algum algoritmo de AM em sua aplicação, uma estratégia de busca foi definida. Kitchenahm e Charters (2007), definem a construção da *string* como a principal atividade da estratégia de busca, sendo indispensável considerar o conjunto de perguntas de pesquisa. Desse modo, a seguir é apresentado os passos para a construção dos termos de busca:

- Os termos de busca foram identificados a partir das perguntas de pesquisa. Os termos foram traduzidos para o inglês, por ter uma literatura mais abrangente;

- Como em algumas literaturas os algoritmos de AM são chamados de técnicas, foi decidido usá-lo como termo;
- Como AM é comumente usado dentro do tópico de IA ou IC, a adição do termo "*Artificial Intelligence*" foi importante para certificar que o uso de diferentes terminologias não iria interferir nos resultados;
- A *string* foi gerada a partir da combinação dos termos. Utilizaram-se aspas duplas, parênteses e operadores lógicos *AND* e *OR*.

Baseado nas definições da tabela 1, os seguintes termos de busca foram construídos:

- A Machine Learning + Customer Relationship Management + Algorithm
- B Machine Learning + Customer Relationship Management + Technique
- C Artificial Intelligence + Customer Relationship Management + Algorithm
- D Artificial Intelligence + Customer Relationship Management + Technique

As estratégias de busca para estudos primários podem ser automática e/ou manual. Este estudo adota a estratégia automática, realizada a partir da execução da *string* de busca. A seguir a tabela 2 mostra a *string* resultante, que foi aplicada em cada base de dado. Portanto, os termos de busca foram concatenados com operadores booleanos *OR* e *AND*.

Tabela 1 – *String* de Busca para a RSL

<i>String</i>
("Machine Learning"AND "Customer Relationship Management"AND "Algorithm") OR ("Machine Learning"AND "Customer Relationship Management"AND "Technique") OR ("Artificial Intelligence"AND "Customer Relationship Management"AND "Algorithm") OR ("Artificial Intelligence"AND "Customer Relationship Management"AND "Technique").

Fonte: Elaborado pela Autora.

3.1.3 Estratégia de Seleção

Para Kitchenham e Charters (2007), a estratégia de seleção deve suprir um conjunto de critérios de inclusão e exclusão e um processo para aplicar tais critérios. Sendo que esses critérios têm como objetivo identificar os estudos que fornecem evidências diretas sobre a questão de pesquisa. Assim, foram especificados critérios de inclusão e exclusão, a fim de que apenas trabalhos relacionados ao contexto fossem selecionados.

3.1.3.1 Critérios de Inclusão

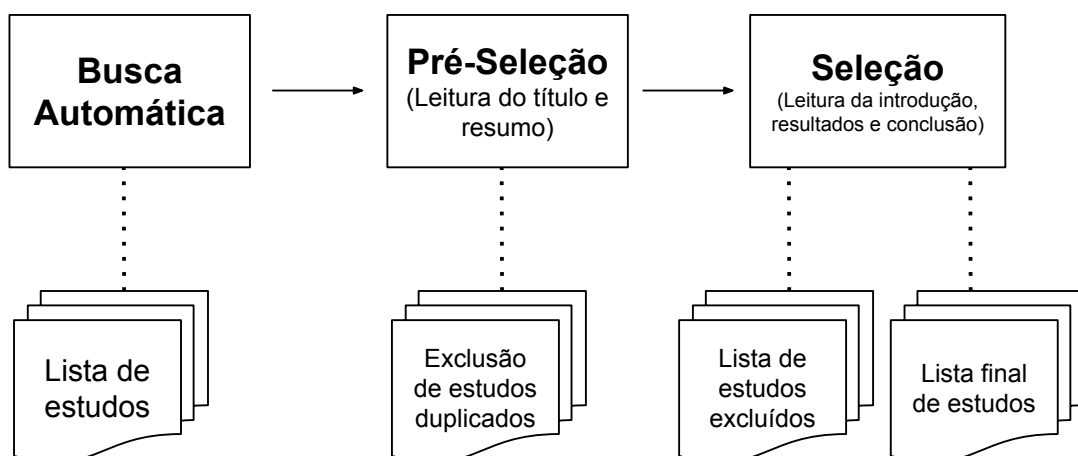
- **CI-1:** artigos que estão no intervalo de tempo escolhido para a análise;
- **CI-2:** artigos que aplicam AM em alguma área de CRM.

3.1.3.2 Critérios de Exclusão

- **CE-1:** o artigo não está disponível eletronicamente na web;
- **CE-2:** o artigo não é apresentado inteiramente no idioma inglês;
- **CE-3:** o artigo foi publicado a mais de seis anos; esse critério foi aplicado assim como em outras revisões de literatura nos campos de IA e AM (CHAI; LIU; NGAI, 2013), (TERÁN; MENSAH; ESTORELLI, 2018).
- **CE-4:** o artigo só está relacionado apenas a AM ou CRM;
- **CE-5:** exclusão de dissertações de mestrado, teses de doutorado, sites, artigos incompletos e artigos de conferência.
- **CE-6:** exclusão de artigos duplicados.

De acordo com Kitchenham (2004), a seleção dos estudos é um processo de multiestágios. Inicialmente, os critérios de seleção devem ser interpretados de forma liberal, a menos que os estudos identificados possam ser claramente excluídos com base em títulos e resumos, cópias completas devem ser obtidas. Assim sendo, a Figura 7 apresenta o processo de seleção dos estudos. A seguir, cada passo é descrito:

Figura 7 – Processo de Seleção dos Estudos.



Fonte: Elaborado pela Autora

Passo 1 - Busca Automática: neste passo, as buscas serão realizadas nas bases de dados eletrônicas selecionadas. Essas bases de dados permitem a filtragem de artigos por ano e por tipo de publicação, dessa forma os critérios de exclusão CE-3 e CE5 poderão ser aplicados. Kitchenham e Charters (2007), afirmam que as buscas iniciais retornam uma grande quantidade de estudos irrelevantes.

Passo 2 - Pré-Seleção: após a execução da busca automática, serão lidos os títulos, resumos e palavras-chaves dos estudos retornados, e estes serão submetidos a análise seguindo os critérios de seleção CI-1, CE-2 e CE-6.

Passo 3 - Seleção: os estudos retornados na fase anterior serão avaliados através da leitura da introdução, resultados e conclusão. Ao final, este processo retornará duas listas, uma com os estudos excluídos e outra com os estudo que respondem às perguntas de pesquisa, e os critérios CI-1 e CE-1 a CE-6, voltam a ser utilizados.

3.1.4 Avaliação de Qualidade do Estudo

A avaliação da qualidade dos estudos é importante para apoiar o processo de inclusão e exclusão e a alocação de ponderação para estudos específicos durante o estágio de síntese de dados. Não há definições universais de qualidade do estudo, mas foi sugerido que a qualidade se relacione com a medida em que o viés é minimizado e a validação externa e interna é maximizada (KITCHENHAM; CHARTERS, 2007); (KHAN et al., 2001).

Para a avaliação dos estudos selecionados na etapa anterior deste trabalho, será desenvolvido um *checklist*, assim como em (SCANNAVINO, 2012). As seguintes questões foram adotadas:

P1. O algoritmo de AM está totalmente definido?

P2. Existe alguma explicação para o uso do algoritmo?

P3. O algoritmo foi validado?

P4. O uso do algoritmo ficou claro?

Para cada pergunta do *checklist*, a seguinte escala será aplicada: Sim (S) - 1 ponto; Não (N) - 0 pontos; Parcialmente (P) - 0.5 ponto. A pontuação total poderá variar entre 0 - 1.0 (muito pobre); 1.1 - 2.0 (razoável); 2.1 - 3.0 (bom) e 3.1 - 4.0 (excelente).

A intenção do *checklist* é classificar os estudos em uma escala de muito pobre, razoável, bom e excelente para ajudar na decisão de inclusão e exclusão dos trabalhos selecionados. Assim, o ponto de corte para excluir estudos da lista será 2.0 (razoável).

3.1.5 Estratégia de Extração

O procedimento de extração é a ponte entre o que é apresentado pelos autores primários e o que é analisado na revisão sistemática.

A estratégia de extração tem por objetivo elaborar o instrumento para a extração dos dados de estudos primários. O instrumento deve ser projetado para coletar as informações necessárias para responder as perguntas de pesquisa (KITCHENHAM; CHARTERS, 2007).

Segundo Silva (2011), a etapa de extração de dados, deve determinar quais dados serão extraídos de cada estudo selecionado e os procedimentos para sua extração. Os dados devem ser definidos e, deve ser criado um formulário.

À vista disso, um formulário será criado a fim de analisar os estudos primários, e responder as questões de pesquisa. Os seguintes dados serão extraídos dos artigos: título do estudo, ano do estudo, país, dimensão do CRM e autores do estudo. Com o objetivo de melhorar a compreensão dos estudos, a análise ocorrerá ao mesmo tempo que a extração.

3.1.6 Síntese e Análise dos Dados

A síntese de dados envolve a comparação e o resumo dos resultados primários incluídos na pesquisa. A síntese pode ser descritiva (qualitativa). No entanto, às vezes é possível complementar uma síntese descritiva com um resumo quantitativo. A utilização de técnicas estatísticas para obter uma síntese quantitativa é referida como meta-análise (KITCHENHAM, 2004). Esta pesquisa tem abordagem qualitativa-quantitativa.

Segundo Creswell (2010), a pesquisa qualitativa envolve, sobretudo, dados baseados em textos e imagens, tem seu relatório final escrito com uma estrutura flexível; bem como seus procedimentos que mesmo nas estratégias, não são uniformes. Desta forma, a pesquisa qualitativa possui estilo indutivo, foco no significado individual e na complexidade de uma situação.

As evidências mapeadas nesta seção serão apresentadas no capítulo 5, assim, servindo de apoio para as respostas para as questões de pesquisa.

4 CONDUÇÃO DA RSL

Este capítulo fornece os resultados da condução da RSL que foram obtidos através da aplicação do protocolo detalhado no capítulo 3.

Kitchenham e Charters (2007), definem que a condução da RSL envolve: identificação da pesquisa, seleção dos estudos, qualidade da avaliação dos estudos, extração dos dados e monitoria dos processos e síntese dos dados.

4.1 Resultados da Seleção

A seleção começou em Março de 2018, então apenas os trabalhos que foram publicados selecionados antes de Março de 2018. Como visto na Figura 7, o processo de seleção tem três etapas, que são apresentadas abaixo:

Passo 1 - Busca Automática: nesta fase, os estudos foram obtidos automaticamente através das seguintes fontes: IEEEExplore, Science Direct, Ebsco e Emerald. A *string* de busca foi executada nas bases de dados que aceitam o recurso. Todas as bases possuíam filtros para aplicar os critérios de exclusão CE-3 e CE-5. Com esse procedimento, 1182 estudos foram retornados ao final da busca, na soma das quatro base de dados, conforme representado na Tabela 2.

Tabela 2 – Resultado da Busca Automática.

Bases de Dados	Estudos Retornados
Ebsco	23
IEEE	75
Science Direct	3306
Emerald	1123
Total	4527

Fonte: Elaborado pela Autora.

Passo 2 - Pré-Seleção: nesta fase, foi encontrado 85 estudos duplicados. Cada um dos 1131 estudos foram analisados pela leitura dos títulos, resumos e palavras-chaves, e foram excluídos aqueles que não estavam de acordo com os critérios de exclusão CE-2 e CE-6.

Passo 3 - Seleção: foi necessário ler a introdução, os resultados e a conclusão de cada estudo. Com isso, o critério de exclusão CE-4 pôde ser aplicado. E, dessa forma, 1060 estudos foram rejeitados. Assim, uma lista foi criada, com os seguintes dados: base de dados, título, autor, resumo e palavras-chave.

Com isso, 100 estudos constavam em mais de uma base de dados e por isso foram excluídos e 4.343 foram rejeitados por não atenderem todos os critérios de seleção. Dessa forma, a lista final de estudos foi obtida, e apenas 1,85% (84) dos estudos se encaixaram nos critérios

de inclusão e exclusão que foram definidos para este trabalho, conforme detalhado na Tabela 3 que retrata o resultado final da seleção.

Tabela 3 – Resultado Final da Seleção.

Estudos Únicos	Estudos Excluídos	Estudos Selecionados
4427	4343	84

Fonte: Elaborado pela Autora.

Os estudos contidos na lista final são os estudos relevantes para este trabalho (Passo 3 - Figura 7). Os mesmos foram baixados e armazenados no serviço de armazenamento e compartilhamento de arquivos em nuvem, para serem utilizados nas próximas etapas da RSL. A Tabela 4, ilustra de outra forma o resultado final da seleção de estudos, relacionando estes com os termos de busca, definidos na seção 3.1.2, onde os números entre parênteses representam a quantidade de estudos que foram selecionados.

Tabela 4 – Total de Artigos Relevantes.

Bases de Dados	Termo de Busca A	Termo de Busca B	Termo de Busca C	Termo de Busca D	Total de Artigos Relevantes
Ebsco	6(0)	7(2)	3(2)	7(1)	5
IEEE	18(3)	15(2)	25(5)	17(3)	13
Science Direct	913(26)	1175(8)	554(14)	664(1)	49
Emerald	213(13)	509(2)	157(1)	244(1)	17
Total					84

Fonte: Elaborado pela Autora.

O apêndice B apresenta a lista dos 84 estudos primários, resultantes da aplicação dos critérios de seleção.

4.2 Avaliação de Qualidade

Os estudos primários que foram selecionados, foram submetidos a uma nova avaliação. Esta com o intuito de classificar os estudos com base em quatro perguntas, referentes ao(s) algoritmo(s) escolhidos, que apoiam a escolha desses estudos e também apresentam um caráter eliminatório, visto que os estudos classificados como razoável serão excluídos.

O apêndice C ilustra a avaliação de qualidade dos 84 estudos primários, retornados na etapa final de seleção de estudos. Esta revela a avaliação de cada estudo em relação a cada pergunta da lista e exibe o total da pontuação.

Como apresentado no apêndice C, o estudo E63 foi excluído. Embora todos os artigos que foram para a última etapa da seleção de estudos, tiveram introdução, resultados e conclusão

lidos, só nesta etapa foi constatado que esse estudo não se encaixa nos critérios de avaliação porque não faz uso de um algoritmo de aprendizado de máquina. Assim, a etapa de avaliação foi muito importante uma vez que a partir das perguntas foi validado a qualidade dos estudos selecionados.

4.3 Resultados da Extração

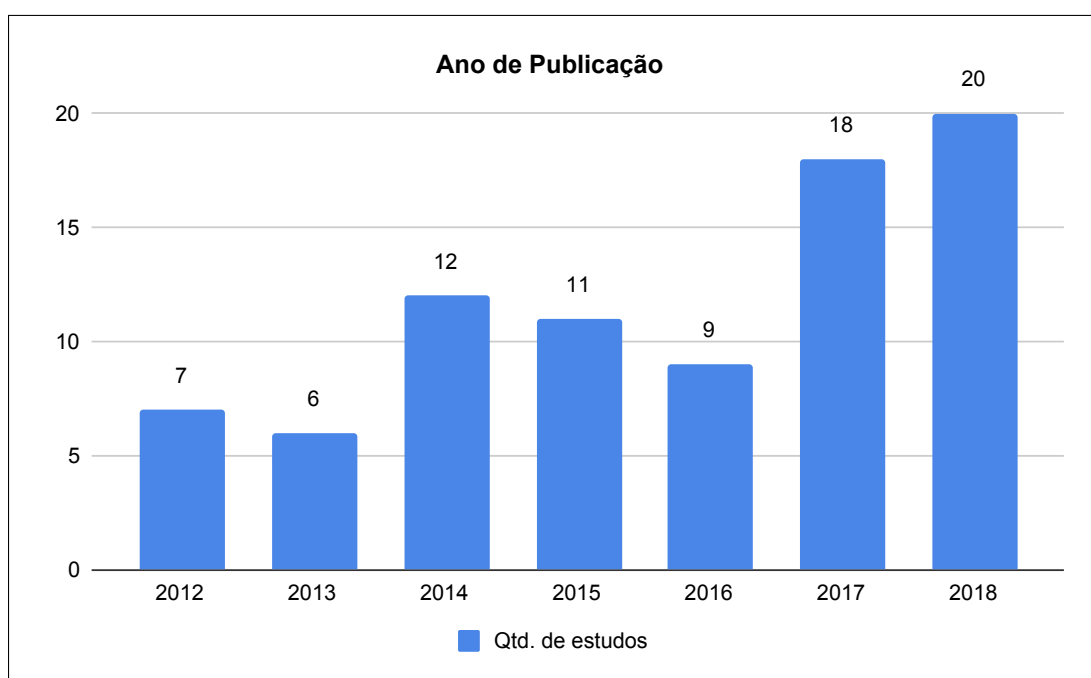
É muito importante saber quais dados devem ser extraídos, esses dados servem para apoiar as respostas das questões de pesquisa. Portanto, uma análise foi realizada com base nas perguntas de pesquisa a fim de que todos os dados extraídos fossem utilizados.

Com o conjunto de estudos primários selecionados e sua qualidade avaliada a atividade de extração de dados foi realizada em todos os 84 artigos.

Como esta etapa foi realizada em conjunto da etapa de seleção de estudos, 4427 estudos foram lidos. Durante este processo, 4343 foram rejeitados por não estarem de acordo com os critérios de exclusão CE-2 e CE-4. E um artigo foi excluído na etapa de avaliação. Então, a extração dos dados aconteceu em 84 estudos primários que foram selecionados. Os seguintes itens foram extraídos de cada estudo: título, ano, país, dimensão do CRM e a referência (a lista completa encontra-se no Apêndice A).

A partir dos dados coletados, foi possível comparar os estudos e analisar as evidências para obter informações relevantes. Deste modo, a Figura 8 ilustra a distribuição temporal dos estudos primários.

Figura 8 – Distribuição dos Estudos por Ano.

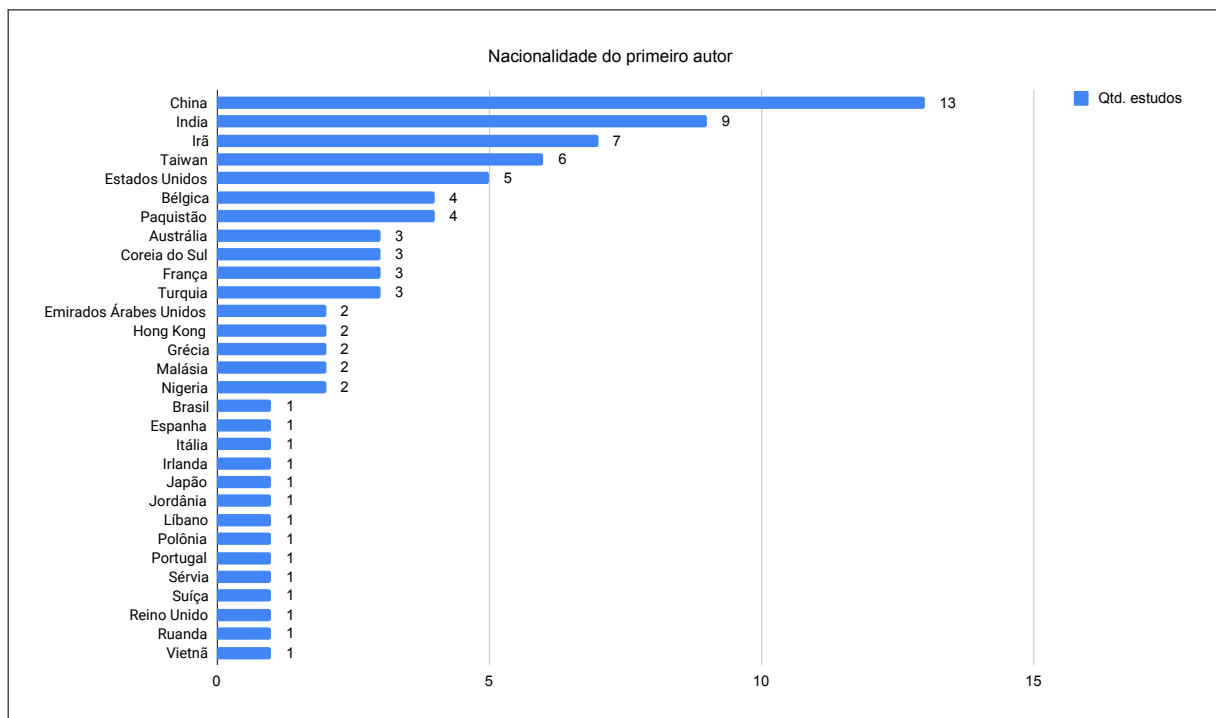


Fonte: Elaborado pela Autora.

A quantidade de publicações sobre aplicações de aprendizado de máquina em CRM era pequena em 2012, teve um pico de 12 artigos em 2014. Nos anos de 2017 e 2018 houve um aumento no número de artigos publicados, respectivamente, 18 e 20 estudos.

Também foi possível analisar quanto à nacionalidade dos estudos primários, para isso levou-se em consideração a origem do primeiro autor (ver Figura 9). Foram 84 autores de 30 países diferentes, foi constatado que 13 (15.47%) dos estudos foram realizados por pesquisadores da China. Seguidos por Índia, Irã e Taiwan com 9, 7 e 6 estudos respectivamente.

Figura 9 – Nacionalidade do Primeiro Autor.



Fonte: Elaborado pela Autora.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A partir do estudo realizado para a área de CRM, considerando uma análise mais abrangente considerando diferentes pontos de vista para delimitar e entender das pesquisas atuais na área três pontos foram discutidos neste trabalho: (i) quais algoritmos de AM são aplicados em processos de CRM; (ii) quais áreas do CRM atualmente tem mais aplicações com AM; (iii) como algoritmos de AM melhoram os processos de CRM.

Este capítulo está organizado para apresentar os resultados da revisão sistemática, que está descrito na seção 5.1.

5.1 Resultados da Revisão Sistemática

A revisão foi conduzida em 5 etapas: (i) identificação de estratégias de busca; (ii) definição de estratégias de seleção de estudos primários; (iii) avaliação dos estudos primários; (iv) extração dos dados encontrados nos estudos selecionados; (v) síntese dos dados dos estudos primários. Esta seção apresenta os resultados alcançados na execução de tais etapas.

5.1.1 QP1. Quais áreas do CRM atualmente tem mais aplicações com AM?

Para responder essa questão de pesquisa, primeiro buscou-se entender quais são as dimensões de CRM e seus elementos. Esta compreensão é apresentada na subseção 2.1.1. Outra necessidade foi delimitar quais fases do AM os estudos primários fazem uso na aplicação do algoritmo, (ver seção 2.5). Com essas duas ideias compreendidas, na etapa de extração dos dados, uma classificação por algoritmos de AM foi realizada em relação aos elementos do CRM.

A Tabela 5 sumariza informações sobre os algoritmos de AM em relação aos elementos de CRM. A primeira coluna apresenta o elemento de CRM, a segunda mostra os algoritmos usados por estudos primários que foram identificados, e a terceira faz referência aos estudos que tem relação com os elementos e com os algoritmos.

Tabela 5 – Classificação dos Algoritmos de AM em Relação aos Elementos de CRM.

Elementos do CRM	Algoritmos	Estudos Primários
Análise de Público-alvo	<i>Neural Network; Multiple Logistic Regression; Weighted Random Forest; RUSBoost; Genetic Algorithm; Decision Tree; Logistic Regression; SVM; Apriori Algorithm; Support Vector Regression</i>	E1, E3, E6, E14, E46, E51, E54, E61, E65, E84

Continua

Elementos do CRM	Algoritmos	Estudos Primários
Segmentação de Mercado	<i>Self-organizing Maps; Support Vector Clustering; XG-Boost; Logistic Regression; Naïve Bayes; Neural Network; C-MK-SVM (Collaborative Multiple Kernel Support Vector Machine); Random Forest; SVM; PCA; Bisection Algorithm; GMDH-SSFS; Fuzzy Cluster Analysis; Decision Tree; k-means; Density-based Algorithm; EM Algorithm; Hierarchical Algorithm</i>	E2, E4, E7, E13, E20, E23, E27, E33, E35, E36, E38, E53, E55, E56, E58, E59, E66, E68, E69, E70, E72, E76, E80, E81, E83
Marketing Direto	<i>Logistic Regression; Projected Gradient Descent Algorithm; Fuzzy; SK-Means</i>	E25, E22, E21, E43, E57
Programa de Fidelidade	<i>Random Forest; Decision Tree; Logistic Regression; Fuzzy; K-NN</i>	E5, E9, E18, E37, E73, E82
Marketing Personalizado	<i>SVM; Exhaustive Algorithm; Genetic Algorithm; Covering Algorithm; Decision Tree; Logistic Regression; Adaptive Boosting; Gradient Boosting; Naïve Bayes; Random Forest; PCA; Neural Network</i>	E12, E19, E26, E29, E40, E44, E50, E49, E62
Monitoramento de Reclamações	<i>Logistic Regression; Random Forest; Gradient Boosting; XG-Boost; AdaBoost, Neural Networks; Extra Trees; Case-based Reasoning; SVM; Naïve Bayes; Decision Tree; NBTree (Naïve Bayes Tree); Co-EM Ensemble; Co-EM SVM; Co-training; Co-EM Bayesian; SMO; Genetic Algorithm</i>	E8, E10, E11, E16, E24, E28, E31, E39, E47, E64, E67, E71, E75, E77, E78, E79
Ciclo de Vida do Cliente	<i>Random Forest; SVM; Naïve Bayes; MK-SVR; Logistic Regression; K-NN; Decision Tree; Weighted K-means</i>	E17, E45, E41, E48, E34

Continua

Elementos do CRM	Algoritmos	Estudos Primários
<i>Up/Cross-selling</i>	<i>MK-STM; Neural Network; Fuzzy</i>	E30, E32, E52, E60
Análise de Afinidade	<i>Neural Network; Logit Model; K-NN; Fuzzy</i>	E15, E42, E74

Fonte: Elaborado pela Autora.

A análise de público-alvo procura identificar, dentro de um grupo, possíveis clientes com base em algumas características semelhantes. Essa análise é altamente recomendada no início do processo de desenvolvimento de um novo produto/serviço, porque há a necessidade de saber pra quem esse produto/serviço é destinado. Como apresentado na tabela acima, a maioria dos algoritmos são do tipo supervisionado, o que não é tão compreensível já que esta é uma tarefa de agrupamento, assim um algoritmo do tipo não supervisionado seria o mais recomendado.

A segmentação de mercado permite identificar as principais diferenças que divide os clientes em grupos que podem ser segmentados. A segmentação de clientes é necessária para obter êxito em várias tarefas a partir dos elementos do CRM, como o *up/cross-selling*, marketing direcionado, fidelização do cliente, etc. 17 estudos foram rotulados como segmentação de mercado, o que representa 20% do total de estudos selecionados, e a maioria dos estudos utilizaram algoritmos de agrupamento.

O Marketing Direto objetiva apresentar uma marca ou empresa para o cliente, de uma forma que a marca fique conhecida. E para isso, aplica estratégias para contactar os potenciais clientes, uma estratégia bastante conhecida é o email marketing. O email marketing é amplamente utilizado pelas empresas, é difícil que alguma pessoa que tenha um email ainda não tenha recebido uma mensagem com ofertas de produtos que foram anteriormente pesquisados. contudo, nesta RSL apenas cinco estudos abordaram essa problemática, e dentre os estudos todos empregaram algoritmos supervisionados em suas aplicações.

O Programa de Fidelidade é uma estratégia empregada pelas empresas para recompensar os clientes que comprem com frequência. Essas recompensas podem ser cupons, produtos com descontos. Apesar de ser uma prática muito conhecida e difundida por empresas, apenas quatro trabalhos se encaixaram nessa temática.

O Marketing personalizado usa de análise de dados para entregar conteúdo personalizado para potenciais clientes ou *leads*. Assim, é possível oferecer melhor experiência pro cliente, aumento da receita, entre outros. Desse modo, na etapa de seleção, nove estudos propuseram alguma aplicação envolvendo AM e marketing personalizado.

O Monitoramento de Reclamações objetiva garantir um excelente atendimento ao cliente. Para isso, as empresas analisam as informações de seus clientes, a fim de melhorar esse processo.

Esse é um tema que desperta o interesse das empresas, uma vez que sabem a proporção negativa que uma reclamação pode causar para os seus produtos ou para sua marca. Foram encontrados dez trabalhos que envolvem esse tópico e AM, sendo que estes apresentam um das maiores variações em termos de tipos de algoritmos utilizados.

Toda compra realizada por um cliente pode ser monitorada antes, durante e pós a compra, ou seja, trata de todas as etapas de um relacionamento empresa-cliente. O objetivo é ajudar na criação de estratégias de *marketing* para melhorar os resultados da empresa. Assim, o ciclo de vida do cliente monitora cinco estágios: segmentação de clientes, aquisição de clientes, rentabilização, retenção de clientes e recuperação de clientes. Contudo, cinco trabalhos foram encontrados e todos utilizaram algoritmos supervisionados.

O *cross-selling*, ou venda cruzada, é a prática de oferecer ao cliente produtos complementares. No entanto, o *upselling*, é uma prática onde as empresas visam fazer o cliente adquirir uma versão melhor de um produto que já está sendo comprado. Para entender melhor, quando um cliente vai em qualquer rede de *fast-food* e compra hambúrguer e batata-frita, o atendente pergunta se este quer outra batata-frita maior ou com algum molho e isso é *upselling*; Quando esse mesmo atendente, perguntar se o cliente quer adicionar refrigerante é *cross-selling*. Apesar de ambas as práticas serem muito comuns nas empresas, pouco está sendo explorado, visto que na RSL realizada, apenas cinco estudos foram encontrados e todos utilizando algoritmos supervisionados.

Análise de Afinidade se baseia na teoria de que, se um cliente compra um determinado item, esse está mais propenso a comprar um outro grupo de itens, ou seja, encontra relações entre itens que não são relacionados. A RSL encontrou apenas dois estudos com esta temática.

Após detalhar todas as evidências encontradas a partir da Tabela 5, foi realizada outra tabela que exprime a quantidade de estudos por elementos e dimensões do CRM (veja Tabela 6).

Tabela 6 – Número de Estudos Primários de Acordo com as Dimensões e Elementos do CRM.

Dimensões	Elementos	Total de Estudos Primários	Nº de Estudos Primários
Identificação do Cliente	Análise de Público-Alvo	10	35
	Segmentação de Mercado	25	
Atração do Cliente	<i>Marketing</i> Direto	5	5
Retenção do Cliente	Programa de Fidelidade	6	31
	<i>Marketing</i> Personalizado	9	
	Monitoramento de Reclamação	16	
Desenvolvimento do Cliente	Ciclo de Vida do Cliente	5	12
	<i>Up/Cross-Selling</i>	4	
	Análise de Afinidade	3	

Fonte: Elaborado pela Autora.

A Tabela 6 apresenta a visão geral das dimensões e elementos em que tais estudos foram aplicados. Os dados demonstram que a maioria dos estudos foram realizados na dimensão de identificação do consumidor. Ao analisar tais trabalhos, foi notado que as aplicações se referem majoritariamente ao elemento de segmentação de mercado. Em seguida, a dimensão de retenção do cliente tem atraído bastante pesquisas. Nesta dimensão, o elemento de monitoramento de reclamações é o que apresenta maior número de aplicações.

Nota-se, também, que o elemento de análise de afinidade é o que apresenta a menor quantidade de estudos primários, com apenas dois. Seguido por *up/cross-selling* e análise de fidelidade. Este último é uma prática bastante conhecida e utilizada por empresas de varejo. Contudo, de acordo com os dados apresentados, pode-se inferir que não há muitos avanços nesta área.

5.1.2 QP2. Quais algoritmos de Aprendizado de Máquina estão sendo aplicados em processos de CRM?

A Tabela 7 apresenta o resumo dos resultados obtidos para responder esta pergunta. Dentre os 83 estudos, foram encontrados 41 tipos de algoritmos, bem como diversos trabalhos utilizaram mais de um algoritmo. Muitos estudos apresentaram algoritmos semissupervisionados e algoritmos novos. Nestes casos, optou-se por apresentá-los na coluna **outros**.

Tabela 7 – Estudos Selecionados e seus Respetivos Algoritmos.

ID	SVM	RNA	LR	AD	RF	KNN	NB	SOM	PCA	AdaBoost	AG	XG-Boost	GB	Fuzzy	Outros
E1		X						X							
E2															
E3		X	X												X
E4															X
E5				X	X										
E6															X
E7												X			X
E8			X							X					
E9			X	X											
E10					X										X
E11		X				X				X		X			X
E12	X	X													
E13			X												
E14											X				
E15		X													
E16														X	
E17	X			X											
E18															X
E19											X				X
E20						X									
E21					X										

Continua

ID	SVM	RNA	LR	AD	RF	KNN	NB	SOM	PCA	AdaBoost	AG	XG-Boost	GB	Fuzzy	Outros
E22								X							X
E23															
E24	X														
E25			X												X
E26			X	X											X
E27							X								
E28			X	X			X		X						
E29	X		X	X	X		X						X		
E30															X
E31															X
E32		X													
E33															X
E34	X				X										
E35															X
E36	X														X
E37														X	
E38														X	
E39															
E40				X											X
E41															
E42		X					X								X
E43														X	

Continua

ID	SVM	RNA	LR	AD	RF	KNN	NB	SOM	PCA	AdaBoost	AG	XG-Boost	GB	Fuzzy	Outros
E44					X	X									
E45			X												
E46	X	X		X											X
E47	X	X		X		X									X
E48															
E49	X		X	X		X									
E50		X	X												
E51			X	X											
E52														X	
E53	X	X	X		X										
E54															X
E55		X													
E56	X	X		X											
E57															X
E58		X						X	X						
E59															X
E60		X													
E61	X	X		X		X		X	X						
E62			X												
E64	X														
E65															X
E66			X												X

Continua

ID	SVM	RNA	LR	AD	RF	KNN	NB	SOM	PCA	AdaBoost	AG	XG-Boost	GB	Fuzzy	Outros
E67			X				X			X					X
E68			X		X										
E69															X
E70															X
E71			X		X										X
E72															
E73							X								
E74														X	X
E75											X				X
E76					X										
E77			X	X	X		X								
E78			X	X											
E79															
E80				X											X
E81								X							X
E82															X
E83															X
E84		X													X

Fonte: Elaborado pela Autora.

Foram identificados 41 algoritmos diferentes nos 83 estudos primários selecionados, nos quais treze foram usados mais de uma vez. A Tabela 8 exibe os algoritmos em relação ao número de estudos primários em que estes foram utilizados, sendo que pode-se destacar que: 15 algoritmos tiveram mais de duas ocorrências; os demais fazem referência aos algoritmos com uma ocorrência; e algoritmos semissupervisionados foram criados para resolver um problema específico.

Deste modo, foi possível constatar que o *Logistic Regression* foi o algoritmo mais utilizado nos estudos abordados por esta revisão, sendo utilizado em 20 estudos primários que correspondem a 23% do total, seguido pelos algoritmos redes neurais e árvore de decisão, utilizados em 40% dos trabalhos.

Tabela 8 – Algoritmos de AM Mais Usados nos Estudos Primários Selecionados.

Algoritmos	Nº de Estudos Primários
<i>Logistic Regression</i>	20
RNA	18
Árvore de Decisão	16
SVM	15
<i>Random Forest</i>	11
K-Means	10
KNN	8
<i>Naïve Bayes</i>	6
SOM	6
<i>Fuzzy</i>	5
AdaBoost	4
PCA	4
Algoritmo Genético	3
<i>XG-Boost</i>	2
<i>Gradient Boosting</i>	2
Outros	37

Fonte: Elaborado pela Autora.

5.1.3 QP3. Como os algoritmos de Aprendizado de Máquina melhoram os processos de CRM?

O CRM sempre utilizou banco de dados para atingir seus objetivos de maneira mais simples. No entanto, até o momento, esse gerenciamento de informações tem se concentrado principalmente na coleta, armazenamento e fornecimento de dados a pedido dos clientes; deixando de lado os processos de extração, análise e interpretação, que dão valor real à informação recebida.

Como os volumes de dados, que as organizações tem que manipular, aumentaram e os tipos de fontes de dados são bem mais amplos, é fácil de entender porque o CRM precisa do suporte da IA, para se adaptar a essa nova era digital. A IA pode ajudar a melhorar os processos

desenvolvidos dentro dos departamentos das organizações, como: *marketing* e *helpdesk*, o que permite responder com mais precisão aos requisitos do cliente. Além disso, pode otimizar estratégias de vendas; consolidar a confiança do cliente; aumentar a satisfação do cliente.

Desse modo, a análise realizada no capítulo anterior, revelou um aumento do uso de AM na pesquisa. No entanto, a maioria dos trabalhos apresenta soluções individuais para problemas específicos ou consiste em exemplos acadêmicos que não são facilmente aplicáveis por outras empresas em geral.

Assim, procurou-se casos de uso de AM em CRM. Martech é um blog que aborda assuntos de AM e IA. Seu estudo fornece *insights* sobre os casos de uso AM em CRM (MARTECH, 2018). A Tabela 9 mostra as áreas e os possíveis valiosos casos de uso de AM para CRM.

Tabela 9 – Áreas e Possíveis Casos de Uso de AM Valiosos para CRM.

Áreas	Casos de Uso de AM
<i>Marketing</i>	• Prever valores de dados de clientes ausentes ou desatualizados • Estratégias de segmentação agrupamento. • Produto/serviço, conteúdo ou recomendação promocional • Melhor canal para participar • Melhor hora para enviar • Modelos para prever as preferências do cliente, o valor da vida útil ou o valor atual do cliente • Otimizar o preço de varejo ou descontos de produtos/serviços • Previsão do tamanho do mercado, receitas/custos de marketing • Planejamento de alocação de orçamento / otimização de mix de mídia • Identificação de conteúdo
Automação de vendas	• Prever <i>leads</i> • Prever atividade de vendas • Contratos de venda • Previsão de vendas e demanda

Continua

Áreas	Casos de Uso de AM
Serviço ao cliente	• Melhor serviço, apoio, ação de treinamento para resolver uma reclamação/problema • Marcar e classificar tickets de suporte • Automatização do processo de suporte • Estimando volumes de serviços, tempos de espera • Otimizar o agendamento e o suporte ao recurso utilizado • Previsão de volume do centro de contato • Modelos para prever níveis de frustração ou satisfação do cliente • Análise de Sentimento com todos os dados do formulário • Modelos para prever a intenção • Visualização dos tópicos de suporte a tendências

Fonte: Elaborado pela Autora.

Todavia, foram selecionados exemplos de artigos que explicitamente declararam as implicações práticas de suas pesquisas. A Tabela 10 apresenta as implicações práticas relacionadas às dimensões de CRM.

Tabela 10 – Implicações Práticas da Literatura e Dimensões do CRM Correspondentes.

ID	Implicação Prática	Dimensão do CRM
E1	Mudança nos padrões de comportamento do cliente para melhorar a eficiência, a satisfação e o equilíbrio do cliente, bem como as vendas rápidas.	Identificação do Cliente
E2	Melhoria no desempenho de segmentação, atingindo a perfeição da regra extraída. Processos de marketing e serviços funcionam de forma mais eficaz.	Identificação do Cliente
E3	Os resultados mostram que o FSR seleciona cerca de 55 % dos recursos inicialmente disponíveis, reduzindo consideravelmente os custos computacionais. Fornece previsão mais precisa do que outras técnicas.	Identificação do Cliente

Continua

ID	Implicação Prática	Dimensão do CRM
E15	Este artigo demonstra o estudo e a simulação de um sistema que pode fornecer uma previsão dos bens necessários para o inventário de uma empresa, dependendo do histórico da venda da empresa em relação aos eventos ocorridos em diferentes períodos.	Desenvolvimento do Cliente
E16	artigo constrói uma ontologia completa de reclamações de clientes. Uma camada geral e superior de ontologia pode ser aplicada a indústrias de serviços orientadas para o cliente face-a-face, como hospitalidade, varejo e indústrias de turismo.	Retenção do Cliente
E17	Com base neste trabalho, as empresas podem fazer estratégias de fornecimento precisas para diferentes categorias de clientes, que visam clientes de alto potencial de várias maneiras, incluindo novos produtos e incentivando um consumo maior.	Desenvolvimento do Cliente
E20	De acordo com os segmentos de mercado apresentados por sua abordagem, as lojas online podem desenvolver marketing promocional e oferecer atendimento personalizado para clientes, que são mais valiosos e promissores.	Atração do Cliente
E24	Os resultados podem ser úteis para direcionar esforços para responder a resenhas on-line em alinhamento com as estratégias (setor hoteleiro), elaborando perfis das avaliações de acordo com o membro e revisando a data.	Retenção do Cliente

Continua

ID	Implicação Prática	Dimensão do CRM
E37	As descobertas fornecem insights aos provedores de serviços móveis sobre como detectar mudanças temporais no comportamento do cliente.	Retenção do Cliente
E54	Este artigo ajuda os gerentes da indústria de fast-food a analisar prontamente o comportamento de seus clientes, a fim de entender suas necessidades e estabelecer relações fortes.	Identificação do Cliente
E55	Os varejistas podem fazer uso do modelo proposto para previsão de demanda de vários itens para melhorar o desempenho do estoque e a rentabilidade das operações.	Identificação do Cliente
E57	Conhecer os sentimentos dos clientes e melhorar os quatro tópicos principais discutidos neste estudo, ou seja, atendimento ao cliente, promoções de bilhetes de LCCs, cancelamentos ou atrasos de voos e gestão pós-reserva ajudarão os LCCs a atrair mais clientes e gerar mais lucros.	Atração do Cliente
E58	Este estudo apresenta a Melhora da precisão para previsão de rotatividade no setor de telecomunicações.	Retenção do Cliente
E59	O resultado da análise do segmento pode fornecer suporte para o design e desenvolvimento de novos produtos, reposicionamento de produtos existentes, desenvolvimento de estratégias de marketing e diferenciação de produtos.	Identificação do Cliente

Continua

ID	Implicação Prática	Dimensão do CRM
E62	A pesquisa enfatiza a importância do aprendizado sensível ao custo para melhorar a gestão de retenção de clientes no contexto de varejo.	Retenção do Cliente

Fonte: Elaborado pela Autora.

A grande maioria dos trabalhos procura combinar ou melhorar os métodos e técnicas de AM para alcançar melhores resultados em comparação com os modelos anteriores, concentrando-se na contribuição dos seus modelos recentemente desenvolvidos. Independentemente do número reduzido de 15 artigos que explicitam claramente as implicações práticas de suas pesquisas, a amostra menor confirma os resultados da Tabela 6, com um número maior de artigos sobre Identificação e Retenção de Clientes.

Apesar da relevância prática da Tabela 10, o número de artigos analisados diminuiu significativamente. Por essa razão, investigamos os sistemas atuais de CRM e sua implementação de algoritmos AM. O GetCRM analisa ferramentas de CRM e funciona como um mercado para essas ferramentas com mais de 50 mil usuários corporativos. Seu recente estudo fornece insights sobre o uso de técnicas de IA e AM para melhorar os processos de CRM e indica oito ferramentas que implementaram AM (YONATAN, 2018).

A Tabela 11 mostra as ferramentas e os processos de CRM aprimorados pela aplicação de algoritmos de AM. A tabela também indica as dimensões do CRM de acordo com as aplicações.

Tabela 11 – Utilização de AM em Ferramentas de CRM.

Ferramentas	Uso de Algoritmos de AM	Dimensão de CRM
Marketo	• Encurta o tempo para a qualificação de <i>leads</i>. • Melhora o engajamento de <i>prospects</i> personalizados.	Identificação de Cliente
Base	• Concentra-se nas vendas, analisando dados e oferecendo feeds internos, bem como recomendações para aumentar as vendas, o acompanhamento de metas e a previsão.	Atração e Desenvolvimento do Cliente

Continua

Ferramentas	Uso de Algoritmos de AM	Dimensão de CRM
Sugar CRM	- Atualiza as informações do cliente automaticamente com base em pontos de dados externos. - Recomenda interações específicas com os clientes.	Identificação e Desenvolvimento do Cliente
Zoho	- Detecta anomalias nas operações de vendas para prever se as equipes atingirão suas metas. - Prevê o tempo ideal para os vendedores alcançarem as perspectivas.	Atração de Cliente
Zendesk	Prevê satisfação analisando a probabilidade de determinados tickets serem avaliados como bons ou ruins.	Retenção de Cliente
Gorgias.io	- Helpdesk para integrar a comunicação com os clientes em diferentes canais. - Cria fluxos de trabalho que priorizam os clientes para problemas específicos.	Retenção de Cliente
Dynamics 365	Fornece tradução, análise de sentimento e identificação de palavras-chave, beneficiando o alcance de vendas e o atendimento social ao cliente	Atração e Retenção de Cliente
Salesforce	- Aprende com dados para prever resultados, prescreve soluções e automatiza tarefas. - Integra dados para gerar pontuação preditiva de <i>leads</i>. - Crie campanhas de <i>marketing</i>.	Identificação de Cliente

Fonte: Elaborado pela Autora.

Como visto, o AM forneceu meios para melhorar os processos e análises no CRM. O AM tem apoiado essas ferramentas para melhorar a automação de processos, previsões, recomendações de vendas e facilitar a comunicação entre empresas e clientes. Parekh (2018), ressalta três etapas que permitem que as empresas melhorem sua experiência com o cliente:

- **Captura de Dados Automatizada:** Como as ferramentas de CRM estão coletando dados

de clientes de vários pontos de contato de integração, a equipe de vendas economiza tempo omitindo o trabalho duplicado.

- **Uso do Produto:** As empresas são capazes de coletar com sucesso os dados que os clientes compartilham rapidamente, gamificando suas ofertas para seus clientes. Quanto mais os clientes usam um produto ou serviço, mais inteligente fica o aprendizado de máquina.
- **Análise de dados:** Os dados precisam ser limpos, classificados e analisados para ajudar uma empresa a melhorar a experiência do cliente.

Apesar de um número maior de estudos sobre identificação e retenção de clientes indicados na revisão de literatura, as ferramentas de CRM selecionadas mostram um foco bem distribuído de algoritmos de AM nas quatro dimensões de CRM. Mostra uma preocupação das ferramentas de CRM para melhorar os processos em todas as dimensões e destaca a lacuna de pesquisa em relação à Atração e Desenvolvimento do Cliente.

Os avanços na pesquisa acadêmica podem influenciar positivamente os desenvolvimentos de curto prazo nas ferramentas existentes, bem como aumentar o número de novas ferramentas. Esta melhoria nos processos de CRM suporta diferentes negócios no aprimoramento do relacionamento com seus clientes e potenciais clientes, influenciando o desempenho de várias empresas em todo o mundo.

6 CONCLUSÕES

Sistemas que envolvem CRM e aprendizado de máquina têm despertado um interesse crescente na comunidade científica. Sendo assim, este trabalho objetivou investigar as lacunas de pesquisa em aplicações de AM relacionadas a processos de CRM.

Considerando as definições apresentadas, nesta dissertação, para as áreas de inteligência computacional e de aprendizado de máquina, a Tabela 12 resume a taxonomia que foi estabelecida para as técnicas consideradas neste trabalho. Apenas os algoritmos com mais de duas ocorrências identificadas nos trabalhos analisados, foram incluídas nesta tabela. A seguir, cada um dos 13 algoritmos de inteligência computacional e/ou aprendizado de máquina são descritos.

- **Lógica fuzzy:** uma forma de lógica multivalorada em que valores de "verdade" de variáveis podem ser qualquer número entre real entre 0 e 1, em contraste com a lógica booleana em que os valores "verdade" de variáveis só podem ser 0 ou 1. A lógica *fuzzy* lida com o conceito de verdade parcial, em que o valor pode variar entre completamente verdadeira e completamente falsa (MAITA, 2016).
- **SVM:** utiliza a teoria do aprendizado estatístico e a teoria da otimização para a classificação binária. Dessa forma, consiste em dado um conjunto de treinamento como entrada, separar o conjunto em duas amostras, estas podendo ser denominadas genericamente como amostras positivas e negativas, assim, gerando duas saídas de classificação (VAPNIK, 2013; NIEUWENHOFF, 2017).
- **SOM:** é uma rede neural não supervisionada, frequentemente usada em tarefas de agrupamento e visualização de dados. O objetivo é encontrar um conjunto de valores de referência e associar cada objeto do conjunto de dados ao vetor de referência mais próximo. O algoritmo depende da inicialização dos vetores de referência. O resultado consiste em um conjunto de vetores de referência que definem implicitamente os grupos (OLIVEIRA et al., 2012 apud KOHONEN, 1990).
- **Naïve Bayes:** realiza a classificação com base nas probabilidades que chegaram, com base no pressuposto de que todas as variáveis são condicionalmente independentes umas das outras. Para estimar os parâmetros (médias e variâncias das variáveis) necessários para a classificação, o classificador requer apenas uma pequena quantidade de dados de treinamento. Também é capaz de lidar com dados reais e discretos. Seu objetivo é encontrar a melhor classe para um documento, maximizando a probabilidade a posteriori conforme a uma equação definida (ELSALAMONY, 2014; OLIVEIRA et al., 2012).
- **KNN:** esse algoritmo armazena os dados de treinamento e quando um novo objeto é submetido para classificação, o algoritmo procura os k registros mais próximos (medida

de distância) desse novo registro. O novo registro é classificado na classe mais comum entre todos os k registros mais próximos (OLIVEIRA et al., 2012).

- **Árvore de Decisão (AD):** funciona como um fluxograma em forma de árvore, em que cada nó (não folha) indica um teste feito sobre um valor. As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence (OLIVEIRA et al., 2012).
- **Algoritmo Genético (AG):** Proposto inicialmente por Holland (1975), é um tipo de algoritmo meta-heurístico de busca e otimização inspirado pelo princípio de seleção natural de Charles Darwin. A ideia central da seleção natural é que o mais forte sobrevive. Através do processo de seleção natural, os organismos se adaptam para otimizar suas chances de sobrevivência em determinado ambiente. Mutações aleatórias ocorrem na descrição genética de um organismo, que é então passado para seus filhos. Se uma mutação se mostrar útil, os filhos terão maior probabilidade de sobreviver para se reproduzir. Se for prejudicial, é menos provável que esses filhos se reproduzam, de modo que a característica ruim irá morrer com elas. Assim, os algoritmos genéticos são naturalmente métodos paralelos e diretos que visam avaliar e modificar conjuntos de soluções simultaneamente (SKIENA, 1998; ZHOU, 2006; POPOV, 2005).
- **Redes Neurais Artificiais (RNA):** são modelos computacionais inspirados no sistema nervoso humano. A arquitetura de uma rede neural artificial define como seus neurônios são organizados em relação um ao outro. Neurônios são unidades computacionais na rede que pesaram os sinais de entrada e produzem um sinal de saída usando uma função de ativação. As principais características da arquitetura de uma rede neural, são: camada de entrada, as camadas ocultas e a camada de saída. A camada de entrada manipula a entrada de dados do ambiente externo. As camadas ocultas são compostas por vários neurônios responsáveis pela extração de padrões. Por fim, a camada de saída é responsável por fornecer a saída final, dependendo dos cálculos feitos antes (NUNES; SILVA, 2018).
- **Logistic Regression (LR):** é uma técnica estatística usada para explorar a relação entre uma variável dependente e pelo menos uma variável independente. Um comparativo com a regressão linear é que esta é usada quando a variável dependente é contínua, enquanto as técnicas de regressão logística são usadas com variáveis dependentes categóricas. A regressão logística, como qualquer outra técnica de modelagem em estatística, visa encontrar o modelo mais adequado e mais econômico e, ao mesmo tempo, mais sensato, para avaliar a relação entre uma variável de resposta e pelo menos uma variável independente (AL-GHAMDI, 2002; MUCHABAIWA, 2013).
- **Principal Componentes Analysis (PCA):** tem como propósito a análise dos dados usados, objetivando sua redução, eliminação de sobreposições e a escolha dos modos mais representativos de dados através de combinações lineares das variáveis originais. Esta

técnica é uma forma de identificar relação entre atributos retirados de dados. E muito útil quando os vetores de atributos possuem muitas dimensões, uma vez que é impossível uma representação gráfica (SAPORETTI, 2016).

- **AdaBoost:** esse algoritmo depende de mudanças contínuas na distribuição de um conjunto de treinamento de maneira iterativa. Em vez de dividir um conjunto de dados de treinamento, vários classificadores são construídos iterativamente a partir do conjunto de dados inteiro. Em cada iteração, o novo classificador de base se concentra em treinar instâncias classificadas incorretamente na iteração anterior. A previsão final do conjunto é feita por votação ponderada, em que a previsão de cada classificador é ponderada de acordo com sua precisão nos conjuntos de dados de treinamento (FREUND; SCHAPIRE; ABE, 1999; FARRASH, 2016).
- **XG-Boost:** é um algoritmo de árvore com aumento de gradiente baseado em algoritmos de aumento de gradiente (*Gradient Boosting*). Em comparação com outras máquinas impulsionadas por gradiente, ele usa uma formalização de modelo mais regularizada para controlar o ajuste excessivo, o que proporciona melhor desempenho. Sendo assim, O *XG-Boost* é um algoritmo de aprendizado supervisionado que implementa um processo chamado *boosting* para gerar modelos precisos. *Boosting* refere-se à técnica de aprendizado conjunto de construir muitos modelos sequencialmente, com cada novo modelo tentando corrigir as deficiências do modelo anterior. Na árvore aumentando cada novo modelo que é adicionado ao conjunto é uma árvore de decisão (FRIEDMAN, 2001; MITCHELL; FRANK, 2017).
- **Random Forest (RF):** o conjunto de dados é dividido aleatoriamente em diversos subconjuntos de tamanho menor. Cada um destes conjuntos é construído por um tipo de amostragem denominado de *bootstrap*, onde cada novo conjunto pode ter algumas amostras incluídas mais de uma vez e outras nenhuma vez. A amostragem *bootstrap* garante que 1/3 das amostras são utilizadas para testar as árvores depois de sua criação. A partir de cada subconjunto reproduzido, uma árvore de decisão é construída. Uma floresta aleatória é uma coleção dessas árvores de decisão. Depois que a floresta está gerada, há muitas árvores de decisão a serem testadas, sendo que todas contribuem para a classificação da amostra, através de um voto sobre qual classe a amostra deve pertencer. A classe mais votada é a atribuída a amostra (BREIMAN, 2001; HAN, 2005; SAPORETTI, 2016).

Tabela 12 – Taxonomia de Algoritmos de Inteligência Computacional e de Aprendizado de Máquina Usadas Neste Trabalho.

Algoritmos	Inteligência computacional	Aprendizado de máquina	
		supervisionado	não-supervisionado
Lógica <i>Fuzzy</i>	x		
SVM		x	
SOM	x		x
<i>Naive Bayes</i>		x	
KNN		x	
Árvore de Decisão		x	
AG	x		
RNA	x	x	
LR		x	
PCA	x		
AdaBoost			x
XG-Boost		x	
<i>Gradient Boosting</i>		x	
RF		x	

Fonte: Elaborado pela Autora.

Algoritmos de AM em CRM são uma tendência emergente na indústria. Com isso, tem atraído a atenção de profissionais e acadêmicos. Este artigo identificou 84 artigos entre 2012 e 2018. Este estudo fornece uma revisão sobre pesquisas usando algoritmos de AM para resolver e/ou automatizar tarefas para CRM. Além disso, fornece uma visão geral dos recentes desenvolvimentos nas atuais ferramentas de CRM. Esta análise fornece uma melhor compreensão das implicações práticas ao usar o AM para melhorar a análise e os processos no CRM.

Estudiosos em todo o mundo têm discutido a aplicação de AM, suas técnicas e algoritmos no campo de CRM. Além disso, as ferramentas de CRM estão constantemente desenvolvendo novas maneiras de analisar os dados do cliente usando AM e IA. A análise mostra que SVM, RNA, *Logistic Regression* e Árvore de Decisão são os algoritmos mais utilizados em AM para lidar com problemas de CRM e melhorar processos. Além disso, identificação e retenção de clientes são as dimensões de CRM com o maior acervo na literatura, com 35 e 31 artigos, respectivamente. No entanto, de 84 artigos, apenas cinco abordam a atração do cliente.

Identificação e retenção de clientes são as dimensões de CRM mais discutidas na literatura, é possível perceber que essas duas dimensões tem um número de estudos primários superior as outras duas. Em relação aos elementos de CRM, a segmentação de clientes é a área com maior número de artigos que aplicam algum algoritmo de AM para melhorar seus processos.

A maioria dos estudos rotulados como identificação do cliente, estão relacionados à Segmentação do Cliente. Esses documentos poderiam fornecer informações para a organização, identificar as preferências e necessidades do cliente e alguns algoritmos de AM, como RNA, SVM, como mostrado na Tabela 5, poderiam ser usadas para atingir esse objetivo. No entanto,

apenas três artigos estão relacionados à Análise de afinidade, que é o elemento de Desenvolvimento do Cliente e podem fornecer às organizações identificar conjuntos de eventos de itens que ocorrem em sequência, por exemplo, compras do produto A têm maior probabilidade de comprar o produto B, se o produto B colocado perto do produto A.

De acordo com as descobertas de pesquisas, acredita-se que mais pesquisas possam ser conduzidas para as dimensões Atração do Cliente e Desenvolvimento do Cliente. A fim de identificar grupos-alvo e manter um relacionamento de longo prazo com os clientes. Assim, as organizações podem reter clientes valiosos.

Além da revisão de literatura, o estudo fornece uma análise das implicações práticas (ver Tabelas 9, 10 e 11) e relaciona estas práticas às dimensões do CRM. Esses resultados são importantes para entender como o AM é capaz de melhorar os processos e quais tarefas na literatura atual do CRM têm sido direcionadas. Além disso, fornece *insights* para empresas que usam o AM para desenvolver soluções de CRM. Essa visão geral pode ajudar pesquisadores e profissionais a entender como o AM pode melhorar as atividades do CRM, bem como as dimensões do CRM menos focadas.

6.1 Limitações e Trabalhos Futuros

Este estudo fornece uma visão geral das aplicações atuais dos algoritmos de AM nos processos de CRM. No entanto, a literatura apenas apontou apenas os benefícios dessas técnicas para os processos de CRM. Portanto, a análise das implicações práticas foi importante para destacar esses benefícios. Uma análise mais profunda poderia validar esses benefícios com especialistas nas áreas de CRM e ICT e fornecer uma explicação melhor sobre como os algoritmos de AM são incorporadas aos sistemas de CRM.

O baixo número de pesquisas em algumas técnicas e dimensões do CRM dificulta uma análise mais profunda. No entanto, indica lacunas de pesquisa para pesquisas futuras.

A busca limitada de bancos de dados por artigos para quatro bancos de dados *on-line* também é uma limitação. Pode haver outros periódicos acadêmicos ou conferências que possam oferecer uma visão mais abrangente da aplicação do AM no CRM.

Os desenvolvimentos recentes na pesquisa e na indústria também apontam para um novo campo chamado “Social CRM”, que visa integrar dados de mídia social aos processos de CRM (ALT; REINHOLD, 2012). Essa integração pode se beneficiar dos avanços no campo do aprendizado de máquina, bem como oferecer novos tópicos para o futuro.

REFERÊNCIAS

- ABDI, F.; KHALILI-DAMGHANI, K.; ABOLMAKAREM, S. Solving customer insurance coverage sales plan problem using a multi-stage data mining approach. *Kybernetes*, Emerald Publishing Limited, 2018.
- AGGARWAL, C. C.; PROCOPIUC, C.; YU, P. S. Finding localized associations in market basket data. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 14, n. 1, p. 51–62, 2002.
- AL-GHAMDI, A. S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*, Elsevier, v. 34, n. 6, p. 729–741, 2002.
- ALBERT, A.; MAASOUMY, M. Predictive segmentation of energy consumers. *Applied energy*, Elsevier, v. 177, p. 435–448, 2016.
- ALBUQUERQUE, P.; ALFINITO, S.; TORRES, C. V. Support vector clustering for customer segmentation on mobile tv service. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 44, n. 6, p. 1453–1464, 2015.
- ALI, M. et al. Prediction of churning behavior of customers in telecom sector using supervised learning techniques. In: IEEE. *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. [S.l.], 2018. p. 1–6.
- ALI, Ö. G.; ARITÜRK, U. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, Elsevier, v. 41, n. 17, p. 7889–7903, 2014.
- ALT, R.; REINHOLD, O. Social-customer-relationship-management (social-crm). *Wirtschaftsinformatik*, Springer, v. 54, n. 5, p. 281–286, 2012.
- AMIN, A. et al. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, Elsevier, v. 237, p. 242–254, 2017.
- AU, W.-H.; CHAN, K. C.; YAO, X. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE transactions on evolutionary computation*, IEEE, v. 7, n. 6, p. 532–545, 2003.
- AYO, F. E. A two-phase multiobjective optimization for benchmarking and evaluating service quality in banks. *International Journal of Intelligent Computing and Cybernetics*, Emerald Publishing Limited, 2018.
- BAHARI, T. F.; ELAYIDOM, M. S. An efficient crm-data mining framework for the prediction of customer behaviour. *Procedia computer science*, Elsevier, v. 46, p. 725–731, 2015.
- BALA, P. K. Improving inventory performance with clustering based demand forecasts. *Journal of Modelling in Management*, Emerald Group Publishing Limited, v. 7, n. 1, p. 23–37, 2012.
- BEGG, R.; PALANISWAMI, M. *Computational intelligence for movement sciences: neural networks and other emerging techniques*. [S.l.]: IGI Global, 2006.
- BERFENFELDT, J. *Customer Relationship Management*. Dissertação (Mestrado) — Luleå University of Technology, Suécia, 2010.

- BHADE, K. et al. A systematic approach to customer segmentation and buyer targeting for profit maximization. In: IEEE. *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. [S.l.], 2018. p. 1–6.
- BOSE, I.; CHEN, X. Detecting the migration of mobile service customers using fuzzy clustering. *Information & Management*, Elsevier, v. 52, n. 2, p. 227–238, 2015.
- BOSE, R. Customer relationship management: key components for it success. *Industrial management & Data systems*, MCB UP Ltd, v. 102, n. 2, p. 89–97, 2002.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BUCZAK, A. L.; GUVEN, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, IEEE, v. 18, n. 2, p. 1153–1176, 2015.
- BULL, C. Strategic issues in customer relationship management (crm) implementation. *Business process management Journal*, MCB UP Ltd, v. 9, n. 5, p. 592–602, 2003.
- CAIGNY, A. D.; COUSSEMENT, K.; BOCK, K. W. D. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, Elsevier, v. 269, n. 2, p. 760–772, 2018.
- CASTANHEIRA, L. G. Aplicação de técnicas de mineração de dados em problemas de classificação de padrões. *Belo Horizonte: UFMG*, 2008.
- CHAGAS, B. N. R. et al. Current applications of machine learning techniques in crm: A literature review and practical implications. In: IEEE. *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. [S.l.], 2018. p. 452–458.
- CHAI, J.; LIU, J. N.; NGAI, E. W. Application of decision-making techniques in supplier selection: A systematic review of literature. *Expert systems with applications*, Elsevier, v. 40, n. 10, p. 3872–3885, 2013.
- CHAPELLE, O.; SCHOLKOPF, B.; ZIEN, A. *Semi-supervised learning*. [S.l.: s.n.], 2006.
- CHEN, Z.-Y.; FAN, Z.-P. Distributed customer behavior prediction using multiplex data: a collaborative mk-svm approach. *Knowledge-Based Systems*, Elsevier, v. 35, p. 111–119, 2012.
- CHEN, Z.-Y.; FAN, Z.-P. Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel svr approach. *Knowledge-Based Systems*, Elsevier, v. 43, p. 123–134, 2013.
- CHEN, Z.-Y.; FAN, Z.-P.; SUN, M. A multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations. *European Journal of Operational Research*, Elsevier, v. 255, n. 1, p. 110–120, 2016.
- CHEUNG, K.-W. et al. Mining customer product ratings for personalized marketing. *Decision Support Systems*, Elsevier, v. 35, n. 2, p. 231–243, 2003.
- CHIANG, W.-Y. Applying data mining for online crm marketing strategy. *British Food Journal*, Emerald Publishing Limited, 2018.

- CHIANG, W.-Y. Identifying high-value airlines customers for strategies of online marketing systems: An empirical case in taiwan. *Kybernetes*, Emerald Publishing Limited, v. 47, n. 3, p. 525–538, 2018.
- CORAZZA, M.; FASANO, G.; MASON, F. An artificial neural network-based technique for on-line hotel booking. *Procedia Economics and Finance*, Elsevier, v. 15, p. 45–55, 2014.
- COURT, D. Getting big impact from big data. *McKinsey Quarterly*, v. 1, n. 1, p. 52–60, 2015.
- COUSSEMENT, K. Improving customer retention management through cost-sensitive learning. *European Journal of Marketing*, Emerald Group Publishing Limited, v. 48, n. 3/4, p. 477–495, 2014.
- CRESWELL, J. W. Projeto de pesquisa métodos qualitativo, quantitativo e misto. In: *Projeto de pesquisa métodos qualitativo, quantitativo e misto*. [S.l.: s.n.], 2010.
- CUI, G.; WONG, M. L.; WAN, X. Targeting high value customers while under resource constraint: partial order constrained optimization with genetic algorithm. *Journal of Interactive Marketing*, Elsevier, v. 29, p. 27–37, 2015.
- DALVI, P. K. et al. Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In: IEEE. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. [S.l.], 2016. p. 1–4.
- DANTAS, E. R. G. et al. O uso da descoberta de conhecimento em base de dados para apoiar a tomada de decisões. *V Simpósio de Excelência em Gestão e Tecnologia-SEGeT*, v. 1, p. 50–60, 2008.
- DEAN, J. *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. [S.l.]: John Wiley & Sons, 2014.
- DELOITTE, L. *The Deloitte Consumer Review. Africa: A 21st century view*. 2016.
- DO, D. et al. Customer churn prediction in an internet service provider. In: IEEE. *2017 IEEE International Conference on Big Data (Big Data)*. [S.l.], 2017. p. 3928–3933.
- DREW, J. H. et al. Targeting customers with statistical and data-mining techniques. *Journal of Service Research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 3, n. 3, p. 205–219, 2001.
- DUCHESSI, P.; LAURÍA, E. J. Decision tree models for profiling ski resorts’ promotional and advertising strategies and the impact on sales. *Expert Systems with Applications*, Elsevier, v. 40, n. 15, p. 5822–5829, 2013.
- D’HAEN, J. et al. Integrating expert knowledge and multilingual web crawling data in a lead qualification system. *Decision Support Systems*, Elsevier, v. 82, p. 69–78, 2016.
- ELSALAMONY, H. A. Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*, Foundation of Computer Science (FCS), v. 85, n. 7, p. 12–22, 2014.
- FACELI, K. et al. *Inteligência artificial: Uma abordagem de aprendizado de máquina*. 2011.
- FAN, Z.-P.; CHE, Y.-J.; CHEN, Z.-Y. Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis. *Journal of Business Research*, Elsevier, v. 74, p. 90–100, 2017.

FARHAT, J.; OWAYJAN, M. Erp neural network inventory control. *Procedia computer science*, Elsevier, v. 114, p. 288–295, 2017.

FARQUAD, M.; BOSE, I. Preprocessing unbalanced data using support vector machine. *Decision Support Systems*, Elsevier, v. 53, n. 1, p. 226–233, 2012.

FARQUAD, M. A. H.; RAVI, V.; RAJU, S. B. Churn prediction using comprehensible support vector machine: An analytical crm application. *Applied Soft Computing*, Elsevier, v. 19, p. 31–40, 2014.

FARRASH, M. *Machine learning ensemble method for discovering knowledge from big data*. Tese (Doutorado) — University of East Anglia, 2016.

FATHIAN, M.; HOSEINPOOR, Y.; MINAEI-BIDGOLI, B. Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*, Emerald Group Publishing Limited, v. 45, n. 5, p. 732–743, 2016.

FAYYAD, U. M. et al. Knowledge discovery and data mining: Towards a unifying framework. In: *KDD*. [S.l.: s.n.], 1996. v. 96, p. 82–88.

FONSECA, J. Metodologia da pesquisa científica. fortaleza: Uec, 2002. apostila. SILVEIRA, DT; CORDOVA, FP *A pesquisa científica*. Cap, v. 2, p. 31–42, 2016.

FREUND, Y.; SCHAPIRE, R.; ABE, N. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, JAPANESE SOC ARTIFICIAL INTELL, v. 14, n. 771-780, p. 1612, 1999.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

GIRAUD-CARRIER, C.; POVEL, O. Characterising data mining software. *Intelligent Data Analysis*, v. 7, p. 181–192, 07 2003.

GOLDBERG, A. B.; ZHU, X. *New directions in semi-supervised learning*. Tese (Doutorado) — University of Wisconsin–Madison, 2010.

GOLDSCHMIDT, R.; PASSOS, E. *Data mining: um guia prático*. [S.l.]: Gulf Professional Publishing, 2005.

GREENBERG, P. *CRM at the speed of light: Capturing and keeping customers in Internet real time*. [S.l.]: Elsevier, 2001.

GRIVA, A. et al. Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, Elsevier, v. 100, p. 1–16, 2018.

HALL, P. et al. An overview of machine learning with sas® enterprise miner™. *SAS Institute Inc*, Citeseer, 2014.

HAN, J. *Data Mining: Concepts and Techniques*. [S.l.: s.n.], 2005.

HAN, J.; KAMBER, M.; PEI, J. Concepts and techniques, second edition. *Morgan Kaufmann*, 2006.

- HANIF, A.; AZHAR, N. Resolving class imbalance and feature selection in customer churn dataset. In: IEEE. *2017 International Conference on Frontiers of Information Technology (FIT)*. [S.l.], 2017. p. 82–86.
- HAYKIN, S. S. et al. *Neural networks and learning machines*. [S.l.]: Pearson education Upper Saddle River, 2009. v. 3.
- HE, Z. et al. Mining action rules from scratch. *Expert Systems with Applications*, Elsevier, v. 29, n. 3, p. 691–699, 2005.
- HOLLAND, J. Adaptation in natural and artificial systems, univ. of mich. press. *Ann Arbor*, 1975.
- HONG, C.-W. Using the taguchi method for effective market segmentation. *Expert systems with applications*, Elsevier, v. 39, n. 5, p. 5451–5459, 2012.
- HONG, T.; KIM, E. Segmenting customers in online stores based on factors that affect the customer's intention to purchase. *Expert Systems with Applications*, Elsevier, v. 39, n. 2, p. 2127–2131, 2012.
- HUANG, Y.; KECHADI, T. An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, Elsevier, v. 40, n. 14, p. 5635–5647, 2013.
- IDOINE, C. et al. Magic quadrant for data science and machine-learning platforms. *Gartner*. [online] Available at: <https://RapidMiner.com/resource/read-gartner-magic-quadrant-data-science-platforms/> [Accessed: 13 Oct. 2018], 2018.
- IDRIS, A.; RIZWAN, M.; KHAN, A. Churn prediction in telecom using random forest and pso based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, Elsevier, v. 38, n. 6, p. 1808–1819, 2012.
- IEEE. Ieee computational intelligence society. *Montreal Computational Intelligence Society*, 2008. Disponível em: <http://ewh.ieee.org/r7/montreal/cis/>.
- JAHROMI, A. T.; STAKHOVYCH, S.; EWING, M. Managing b2b customer churn, retention and profitability. *Industrial Marketing Management*, Elsevier, v. 43, n. 7, p. 1258–1268, 2014.
- JAMES, G. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112.
- JASPERS, S.; TROYER, E. D.; AERTS, M. Machine learning techniques for the automation of literature reviews and systematic reviews in efsa. *EFSA Supporting Publications*, Wiley Online Library, v. 15, n. 6, p. 1427E, 2018.
- JIANG, S. et al. Durable product review mining for customer segmentation. *Kybernetes*, Emerald Group Publishing Limited, v. 44, n. 1, p. 124–138, 2015.
- JIANG, T.; TUZHILIN, A. Segmenting customers from population to individuals: Does 1-to-1 keep your customers forever? *IEEE Transactions on knowledge and data engineering*, IEEE, v. 18, n. 10, p. 1297–1311, 2006.
- JOHN, G. H. *Enhancements to the data mining process*. Tese (Doutorado) — stanford university Ph. D. thesis, 1997.
- KERAMATI, A. et al. Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, Elsevier, v. 24, p. 994–1012, 2014.

KHALILI-DAMGHANI, K.; ABDI, F.; ABOLMAKAREM, S. Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. *Applied Soft Computing*, Elsevier, v. 73, p. 816–828, 2018.

KHAN, K. S. et al. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*. [S.l.]: NHS Centre for Reviews and Dissemination, 2001.

KHODABANDEHLOU, S.; RAHMAN, M. Z. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, Emerald Publishing Limited, v. 19, n. 1/2, p. 65–93, 2017.

KHODAKARAMI, F.; CHAN, Y. E. Exploring the role of customer relationship management (crm) systems in customer knowledge creation. *Information & Management*, Elsevier, v. 51, n. 1, p. 27–42, 2014.

KIM, J.; KANG, P. Late payment prediction models for fair allocation of customer contact lists to call center agents. *Decision Support Systems*, Elsevier, v. 85, p. 84–101, 2016.

KIM, K.; JUN, C.-H.; LEE, J. Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Systems with Applications*, Elsevier, v. 41, n. 15, p. 6575–6584, 2014.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Citeseer, 2007.

KOHONEN, T. The self-organizing map. *Proceedings of the IEEE*, IEEE, v. 78, n. 9, p. 1464–1480, 1990.

KRACKLAUER, A. H.; MILLS, D. Q.; SEIFERT, D. Customer management as the origin of collaborative customer relationship management. In: *Collaborative Customer Relationship Management*. [S.l.]: Springer, 2004. p. 3–6.

KUBAT, M. et al. Itemset trees for targeted association querying. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 15, n. 6, p. 1522–1534, 2003.

LATHAUWER, L. D.; MOOR, B. D.; VANDEWALLE, J. Blind source separation by higher-order singular value decomposition. In: *Proc. EUSIPCO*. [S.l.: s.n.], 1994. v. 1, p. 175–178.

LEE, C.-H.; WANG, Y.-H.; TRAPPEY, A. J. Ontology-based reasoning for the intelligent handling of customer complaints. *Computers & Industrial Engineering*, Elsevier, v. 84, p. 144–155, 2015.

LEMON, K. N.; VERHOEF, P. C. Understanding customer experience throughout the customer journey. *Journal of marketing*, American Marketing Association, v. 80, n. 6, p. 69–96, 2016.

LI, H. et al. Supervised massive data analysis for telecommunication customer churn prediction. In: IEEE. *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*. [S.l.], 2016. p. 163–169.

- LIAO, S.-H.; CHEN, Y.-J. Mining customer knowledge for electronic catalog marketing. *Expert Systems with Applications*, Elsevier, v. 27, n. 4, p. 521–532, 2004.
- LIAU, B. Y.; TAN, P. P. Gaining customer knowledge in low cost airlines through text mining. *Industrial management & data systems*, Emerald Group Publishing Limited, v. 114, n. 9, p. 1344–1359, 2014.
- LIN, W.-C.; TSAI, C.-F.; KE, S.-W. Dimensionality and data reduction in telecom churn prediction. *Kybernetes*, Emerald Group Publishing Limited, v. 43, n. 5, p. 737–749, 2014.
- LING, R.; YEN, D. C. Customer relationship management: An analysis framework and implementation strategies. *Journal of computer information systems*, Taylor & Francis, v. 41, n. 3, p. 82–97, 2001.
- LIU, D. User satisfaction management in e-government: One k-means algorithm-based analysis. In: IEEE. *2018 5th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS)*. [S.l.], 2018. p. 1–5.
- LIU, N. et al. Handling class imbalance in customer behavior prediction. In: IEEE. *2014 International Conference on Collaboration Technologies and Systems (CTS)*. [S.l.], 2014. p. 100–103.
- LU, N. et al. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, IEEE, v. 10, n. 2, p. 1659–1665, 2014.
- MADEO, R. C. et al. A committee machine implementing the pattern recognition module for fingerspelling applications. In: ACM. *Proceedings of the 2010 ACM Symposium on Applied Computing*. [S.l.], 2010. p. 954–958.
- MAITA, A. C. Um estudo da aplicação de técnicas de inteligência computacional e de aprendizado em máquina de mineração de processos de negócio. *São Paulo*, 2016.
- MAŁECKI, K.; WĄTRÓBSKI, J. The classification of internet shop customers based on the cluster analysis and graph cellular automata. *Procedia computer science*, Elsevier, v. 112, p. 2280–2289, 2017.
- MARTECH. *Machine Learning & AI Use Cases for CRM*. 2018. Disponível em: <http://vincejeffs.com/ai-use-cases-crm>.
- MARTINS, I. M. A mineração de dados para descoberta de conhecimento e uma oferta adequada no canal de televisão aberta. 2010.
- MAU, S.; PLETIKOSA, I.; WAGNER, J. Forecasting the next likely purchase events of insurance customers. *International Journal of Bank Marketing*, Emerald Publishing Limited, 2018.
- MEIRE, M.; BALLINGS, M.; POEL, D. Van den. The added value of auxiliary data in sentiment analysis of facebook posts. *Decision Support Systems*, Elsevier, v. 89, p. 98–112, 2016.
- MEIRE, M.; BALLINGS, M.; POEL, D. Van den. The added value of social media data in b2b customer acquisition systems: A real-life experiment. *Decision Support Systems*, Elsevier, v. 104, p. 26–37, 2017.

- MILOŠEVIĆ, M.; ŽIVIĆ, N.; ANDJELKOVIĆ, I. Early churn prediction with personalized targeting in mobile social games. *Expert Systems with Applications*, Elsevier, v. 83, p. 326–332, 2017.
- MISHRA, K.; RANI, R. G. *A Machine Learning Approach for Churn Prediction in Telecommunication*. Tese (Doutorado), 2017.
- MITCHELL, R.; FRANK, E. Accelerating the xgboost algorithm using gpu computing. *PeerJ Computer Science*, PeerJ Inc., v. 3, p. e127, 2017.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of machine learning*. [S.l.]: MIT press, 2012.
- MOMTAZ, N. J.; ALIZADEH, S.; VAGHEFI, M. S. A new model for assessment fast food customer behavior case study: An iranian fast-food restaurant. *British food journal*, Emerald Group Publishing Limited, v. 115, n. 4, p. 601–613, 2013.
- MONTGOMERY, L. et al. Customer support ticket escalation prediction using feature engineering. *Requirements Engineering*, Springer, v. 23, n. 3, p. 333–355, 2018.
- MORO, S.; RITA, P.; COELHO, J. Stripping customers' feedback on hotels through data mining: The case of las vegas strip. *Tourism management perspectives*, Elsevier, v. 23, p. 41–52, 2017.
- MUCHABAIWA, H. *Logistic regression to determine significant factors associated with share price change*. Tese (Doutorado), 2013.
- NGAI, E. W.; XIU, L.; CHAU, D. C. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, Elsevier, v. 36, n. 2, p. 2592–2602, 2009.
- NIEUWENHOFF, N. *Uma comparação da aplicação de métodos computacionais de classificação de dados aplicados ao consumo de cinema no Brasil*. Tese (Doutorado) — Universidade de São Paulo, 2017.
- NUNES, I.; SILVA, H. S. D. *Artificial neural networks: A practical course*. [S.l.]: Springer, 2018.
- OGWUELEKA, F. N. et al. Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank. *Human factors and ergonomics in manufacturing & service industries*, Wiley Online Library, v. 25, n. 1, p. 28–42, 2015.
- OLIVEIRA, V. L. M. et al. Analytical customer relationship management in retailing supported by data mining techniques. 2012.
- ONG, Y.-S.; LIM, M. H.; CHEN, X. Research frontier-memetic computation—past, present & future. *IEEE Computational Intelligence Magazine*, v. 5, n. 2, p. 24, 2010.
- OTIENO, B.; WILLIAMS, N. J.; MCSHARRY, P. Customer segmentation for east african microgrid consumers. In: IEEE. *2018 IEEE PES/IAS PowerAfrica*. [S.l.], 2018. p. 468–473.
- PAN, S. et al. Using customer-related data to enhance e-grocery home delivery. *Industrial Management & Data Systems*, Emerald Publishing Limited, v. 117, n. 9, p. 1917–1933, 2017.

- PAREKH, L. *Why Artificial Intelligence is the Next Big Thing in CRM*. 2018. Disponível em: <<https://inc42.com/resources/why-artificial-intelligence-is-the-next-big-thing-in-crm>>.
- PAYNE, A.; FROW, P. A strategic framework for customer relationship management. *Journal of marketing*, SAGE Publications Sage CA: Los Angeles, CA, v. 69, n. 4, p. 167–176, 2005.
- PEELEN, E. *Customer Relationship Management*. [S.l.]: Financial Times/ Prentice Hall, 2005.
- PEPPERS, D.; ROGERS, M. to 1—um guia executivo para entender e implantar estratégias de customer relationship management. *São Paulo, Jan*, 2000.
- PÉREZ, J. M. *Inteligencia computacional inspirada en la vida*. [S.l.]: Servicio Publicaciones UMA, 2010. v. 36.
- POPOV, A. Genetic algorithms for optimization. *User Manual, Hamburg*, v. 2013, 2005.
- POTTALA, M. et al. Artificial intelligence: Artificial intelligence in sports. Centria-ammattikorkeakoulu, 2018.
- PRINZIE, A.; POEL, D. Van den. Constrained optimization of data-mining problems to improve model performance: A direct-marketing application. *Expert Systems with Applications*, Elsevier, v. 29, n. 3, p. 630–640, 2005.
- QADADEH, W.; ABDALLAH, S. Customers segmentation in the insurance company (tic) dataset. *Procedia computer science*, Elsevier, v. 144, p. 277–290, 2018.
- QUIJANO-SANCHEZ, L.; LIBERATORE, F. The big chase: A decision support system for client acquisition applied to financial networks. *Decision Support Systems*, Elsevier, v. 98, p. 49–58, 2017.
- RANDOLPH, J. J. A guide to writing the dissertation literature review. *Practical assessment, research & evaluation*, v. 14, n. 13, p. 1–13, 2009.
- REYNOLDS, J. *A practical guide to CRM: Building more profitable customer relationships*. [S.l.]: CRC Press, 2002.
- REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. [S.l.]: Editora Manole Ltda, 2003.
- RODAN, A.; FARIS, H. Echo state network with svm-readout for customer churn prediction. In: IEEE. *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. [S.l.], 2015. p. 1–5.
- RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. [S.l.]: Malaysia; Pearson Education Limited,, 2016.
- SADEK, A. W. Artificial intelligence applications in transportation. *Transportation Research Circular*, Citeseer, p. 1–7, 2007.
- SAMUEL, A. *Some studies in machine learning using the game of checkers*. Reprinted in EA Feigenbaum & J. Feldman (Eds.)(1963). *Computers and thought*. [S.l.]: McGraw-Hill, 1959.
- SANTOS, H. G. d. *Comparação da performance de algoritmos de machine learning para a análise preditiva em saúde pública e medicina*. Tese (Doutorado) — Universidade de São Paulo, 2018.

- SAPORETTI, C. M. *Comparação de técnicas de inteligência computacional para classificação de dados petrográficos*. Dissertação (Mestrado), 2016.
- SARVARI, P. A.; USTUNDAG, A.; TAKCI, H. Performance evaluation of different customer segmentation approaches based on rfm and demographics analysis. *Kybernetes*, Emerald Group Publishing Limited, v. 45, n. 7, p. 1129–1157, 2016.
- SCANNAVINO, K. R. F. *Evidence-based software engineering: systematic literature review process based on visual text mining*. Tese (Doutorado) — Universidade de São Paulo, 2012.
- SEYFIOĞLU, M. S.; DEMIREZEN, M. U. A hierarchical approach for sentiment analysis and categorization of turkish written customer relationship management data. In: IEEE. *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. [S.l.], 2017. p. 361–365.
- SILVA, M. A. G. *Processo de revisão sistemática*. Dissertação (Mestrado) — UNIVERSIDADE DE SÃO PAULO, Brasil, 2011.
- SKIENA, S. S. *The algorithm design manual: Text*. [S.l.]: Springer Science & Business Media, 1998. v. 1.
- SOROUGH, A.; BAHREININEJAD, A.; BERG, J. van den. A hybrid customer prediction system based on multiple forward stepwise logistic regression mode. *Intelligent Data Analysis*, IOS Press, v. 16, n. 2, p. 265–278, 2012.
- STUART, R.; PETER, N. et al. *Artificial intelligence: a modern approach*. [S.l.]: Prentice Hall Upper Saddle River, NJ, USA:, 2003.
- SUMATHI, S.; SIVANANDAM, S. *Introduction to data mining and its applications*. [S.l.]: Springer, 2006. v. 29.
- SUN, Y. et al. Analyzing high speed rail passengers' train choices based on new online booking data in china. *Transportation Research Part C: Emerging Technologies*, Elsevier, v. 97, p. 96–113, 2018.
- SWIFT, R. S. *Accelerating customer relationships: Using CRM and relationship technologies*. [S.l.]: Prentice Hall Professional, 2001.
- TAMPELINI, L. G. et al. Uma aplicação de redes de elman no tratamento e predição de sTM Eries temporais hidrolTM Ogicas. In: *Anais do 9 Congresso Brasileiro de Redes Neurais*. Ouro Preto, MG: SBRN, 2009. p. 1–5.
- TANAKA, T. et al. Classifying and understanding prospective customers via heterogeneity of supermarket stores. *Procedia Computer Science*, Elsevier, v. 112, p. 956–964, 2017.
- TERÁN, L.; MENSAH, A. O.; ESTORELLI, A. A literature review for recommender systems techniques used in microblogs. *Expert Systems with Applications*, Elsevier, v. 103, p. 63–73, 2018.
- THULIN, J. *Machine learning-based classifiers in the Direkt Profil grammatical profiling system*. Dissertação (Mestrado), 2007.
- TRIVEDI, S. K.; DEY, S.; KUMAR, A. Capturing user sentiments for online indian movie reviews. Emerald Insight, 2018.

- TSE, Y. K. et al. Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward tesco. *Industrial Management & Data Systems*, Emerald Group Publishing Limited, v. 116, n. 6, p. 1178–1200, 2016.
- VAFEIADIS, T. et al. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, Elsevier, v. 55, p. 1–9, 2015.
- VÄHÄKAINU, P.; NEITTAANMÄKI, P. Tekoäly terveydenhuollossa. *Informaatioteknologian tiedekunnan julkaisuja/Jyväskylän yliopisto*, Jyväskylän yliopisto, n. 2018, 45, 2018.
- VALENTIN, L. G. *Uma arquitetura de software para sistemas de descoberta de conhecimento em banco de dados*. Tese (Doutorado) — Dissertação (Mestrado em Ciência da Computação)—Universidade Estadual de Maringá, 2006.
- VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 2013.
- VERBRAKEN, T. et al. Predicting online channel acceptance with social network data. *Decision Support Systems*, Elsevier, v. 63, p. 104–114, 2014.
- VYAS, R. et al. Predicting inactiveness in telecom (prepaid) sector: A complex bigdata application. In: IEEE. *2018 International Conference on Information Technology (ICIT)*. [S.l.], 2018. p. 39–43.
- WANG, L. et al. On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems*, Elsevier, v. 76, p. 3–13, 2015.
- WANG, Y.; LU, X.; TAN, Y. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electronic Commerce Research and Applications*, Elsevier, v. 29, p. 1–11, 2018.
- WENG, C.-H. Revenue prediction by mining frequent itemsets with customer analysis. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 63, p. 85–97, 2017.
- WENG, Y.; YU, J.; RAJAGOPAL, R. Probabilistic baseline estimation based on load patterns for better residential customer rewards. *International Journal of Electrical Power & Energy Systems*, Elsevier, v. 100, p. 508–516, 2018.
- WOO, J. Y.; BAE, S. M.; PARK, S. C. Visualization method for customer targeting using customer map. *Expert Systems with Applications*, Elsevier, v. 28, n. 4, p. 763–772, 2005.
- XIAO, J. et al. Gmdh-based semi-supervised feature selection for customer classification. *Knowledge-Based Systems*, Elsevier, v. 132, p. 236–248, 2017.
- XU, Y. et al. Adopting customer relationship management technology. *Industrial management & data systems*, MCB UP Ltd, v. 102, n. 8, p. 442–452, 2002.
- YADEGARIDEHKORDI, E. et al. Predicting determinants of hotel success and development using structural equation modelling (sem)-anfis method. *Tourism Management*, Elsevier, v. 66, p. 364–386, 2018.
- YANG, Y. et al. An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications. *Knowledge-Based Systems*, Elsevier, v. 162, p. 202–210, 2018.

- YIN, C. et al. A new recommendation system on the basis of consumer initiative decision based on an associative classification approach. *Industrial Management & Data Systems*, Emerald Publishing Limited, v. 118, n. 1, p. 188–203, 2018.
- YONATAN, R. 8 *Machine Learning CRMs That Are Moving The Industry Forward*. GetCRM, 2018. Disponível em: <<https://getcrm.com/blog/machine-learning-crm/>>.
- YOU, Z. et al. A decision-making framework for precision marketing. *Expert Systems with Applications*, Elsevier, v. 42, n. 7, p. 3357–3367, 2015.
- ZHANG, Z. et al. A hybrid fuzzy-based personalized recommender system for telecom products/services. *Information Sciences*, Elsevier, v. 235, p. 117–129, 2013.
- ZHENG, X.; ZHU, S.; LIN, Z. Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. *Decision Support Systems*, Elsevier, v. 56, p. 211–222, 2013.
- ZHOU, Y. Study on genetic algorithm improvement and application. Worcester Polytechnic Institute, 2006.
- ZHU, X.; LAFFERTY, J.; ROSENFELD, R. *Semi-supervised learning with graphs*. Tese (Doutorado) — Carnegie Mellon University, language technologies institute, school of . . . , 2005.
- ZUYLEN, H. van. Difference between artificial intelligence and traditional methods. *Artificial Intelligence Applications to Critical Transportation Issues*, p. 3, 2012.

A EXTRAÇÃO DOS DADOS DOS ESTUDOS PRIMÁRIOS SELECIONADOS

A seguir são apresentados os dados dos 64 estudos primários selecionados na etapa de extração desta RSL.

Tabela 13 – Lista Final dos Estudos Primários .

ID	Título	Ano	País	Dimensão	Referência
E1	<i>Neural Network and Classification Approach in Identifying Customer Behavior in the Banking Sector: A Case Study of an International Bank</i>	2015	Nigéria	Retenção do Cliente	(OGWUELEKA et al., 2015)
E2	<i>Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis</i>	2016	Turquia	Identificação do Cliente	(SARVARI; USTUNDAG; TAKCI, 2016)
E3	<i>A hybrid customer prediction system based on multiple forward stepwise logistic regression mode</i>	2012	Irã	Identificação do Cliente	(SOROUSH; BAHREINI-NEJAD; BERG, 2012)
E4	<i>Support vector clustering for customer segmentation on mobile tv service</i>	2015	Brasil	Identificação do Cliente	(ALBUQUERQUE; ALFIRENITO; TORRES, 2015)
E5	<i>Supervised massive data analysis for telecommunication customer churn prediction</i>	2016	China	Retenção do Cliente	(LI et al., 2016)
E6	<i>Handling class imbalance in customer behavior prediction</i>	2014	Emirados Árabes Unidos	Identificação do Cliente	(LIU et al., 2014)
E7	<i>A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data</i>	2017	Turquia	Identificação do Cliente	(SEYFIOĞLU; DEMIREZEN, 2017)
E8	<i>A customer churn prediction model in telecom industry using boosting</i>	2014	Austrália	Retenção do Cliente	(LU et al., 2014)
E9	<i>Analysis of customer churn prediction in telecom industry using decision trees and logistic regression</i>	2016	Índia	Retenção do Cliente	(DALVI et al., 2016)
E10	<i>Resolving class imbalance and feature selection in customer churn dataset</i>	2017	Paquistão	Retenção do Cliente	(HANIF; AZHAR, 2017)
E11	<i>Customer churn prediction in an internet service provider</i>	2017	Vietnã	Retenção do Cliente	(DO et al., 2017)
E12	<i>Echo state network with SVM-readout for customer churn prediction</i>	2015	Jordânia	Retenção do Cliente	(RODAN; FARIS, 2015)
E13	<i>Classifying and understanding prospective customers via heterogeneity of supermarket Stores</i>	2017	Japão	Identificação do Cliente	(TANAKA et al., 2017)

Continua

ID	Título	Ano	País	Dimensão	Referência
E14	Targeting high value customers while under resource constraint: partial order constrained optimization with genetic algorithm	2015	Hong Kong	Identificação do Cliente	(CUI; WONG; WAN, 2015)
E15	ERP Neural Network Inventory Control	2017	Líbano	Desenvolvimento do Cliente	(FARHAT; OWAYJIAN, 2017)
E16	Ontology-based reasoning for the intelligent handling of customer complaints	2015	Taiwan	Retenção do Cliente	(LEE; WANG; TRAPPEY, 2015)
E17	A decision-making framework for precision marketing	2015	China	Desenvolvimento do Cliente	(YOU et al., 2015)
E18	Revenue prediction by mining frequent itemsets with customer analysis	2017	Taiwan	Retenção do Cliente	(WENG, 2017)
E19	Customer churn prediction in the telecommunication sector using a rough set approach	2017	Paquistão	Retenção do Cliente	(AMIN et al., 2017)
E20	Segmenting customers in online stores based on factors that affect the customer's intention to purchase	2012	Coreia do Sul	Identificação do Cliente	(HONG; KIM, 2012)
E21	The added value of social media data in B2B customer acquisition systems: A real-life experiment	2017	Bélgica	Atração do Cliente	(MEIRE; BALLINGS; POEL, 2017)
E22	The BIG CHASE: A decision support system for client acquisition applied to financial networks	2017	Espanha	Atração do Cliente	(QUIJANO-SANCHEZ; LIBERATORE, 2017)
E23	Using the Taguchi method for effective market segmentation	2012	Taiwan	Identificação do Cliente	(HONG, 2012)
E24	Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip	2017	Portugal	Retenção do Cliente	(MORO; RITA; COELHO, 2017)
E25	Integrating expert knowledge and multilingual web crawling data in a lead qualification system	2016	Bélgica	Atração do Cliente	(D'HAEN et al., 2016)
E26	Managing B2B customer churn, retention and profitability	2014	Austrália	Retenção do Cliente	(JAHROMI; STAKHOVYCH; EWING, 2014)

Continua

ID	Título	Ano	País	Dimensão	Referência
E27	<i>An efficient CRM-data mining framework for the prediction of customer behaviour</i>	2015	Índia	Identificação do Cliente	(BAHARI; ELAYIDOM, 2015)
E28	<i>A comparison of machine learning techniques for customer churn prediction</i>	2015	Grécia	Retenção do Cliente	(VAFEIADIS et al., 2015)
E29	<i>Early churn prediction with personalized targeting in mobile social games</i>	2017	Sérvia	Retenção do Cliente	(MILOŠEVIĆ; ŽIVIĆ; ANDJELKOVIĆ, 2017)
E30	<i>A multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations</i>	2016	China	Desenvolvimento do Cliente	(CHEN; FAN; SUN, 2016)
E31	<i>Churn prediction using comprehensible support vector machine: An analytical CRM application</i>	2014	Índia	Retenção do Cliente	(FARQUAD; RAVI; RAJU, 2014)
E32	<i>An Artificial Neural Network-based technique for on-line hotel booking</i>	2014	Itália	Desenvolvimento do Cliente	(CORAZZA; FASANO; MASON, 2014)
E33	<i>Predictive segmentation of energy consumers</i>	2016	Estados Unidos	Identificação do Cliente	(ALBERT; MAASOUMY, 2016)
E34	<i>The added value of auxiliary data in sentiment analysis of Facebook posts</i>	2016	Bélgica	Desenvolvimento do Cliente	(MEIRE; BALLINGS; POEL, 2016)
E35	<i>GMDH-based semi-supervised feature selection for customer classification</i>	2017	China	Identificação do Cliente	(XIAO et al., 2017)
E36	<i>Distributed customer behavior prediction using multiplex data: a collaborative MK-SVM approach</i>	2012	China	Identificação do Cliente	(CHEN; FAN, 2012)
E37	<i>Detecting the migration of mobile service customers using fuzzy clustering</i>	2015	Índia	Retenção do Cliente	(BOSE; CHEN, 2015)
E38	<i>The classification of internet shop customers based on the cluster analysis and graph cellular automata</i>	2017	Polónia	Identificação do Cliente	(MAŁECKI; WATRÓBSKI, 2017)
E39	<i>Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach</i>	2013	China	Retenção do Cliente	(ZHENG; ZHU; LIN, 2013)

Continua

ID	Título	Ano	País	Dimensão	Referência
E40	<i>Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales</i>	2013	Estados Unidos	Retenção do Cliente	(DUCHESSI; LAURÍA, 2013)
E41	<i>Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis</i>	2017	China	Desenvolvimento do Cliente	(FAN; CHE; CHEN, 2017)
E42	<i>On the brink: Predicting business failure with mobile location-based checkins</i>	2015	Estados Unidos	Desenvolvimento do Cliente	(WANG et al., 2015)
E43	<i>Predicting determinants of hotel success and development using Structural Equation Modelling (SEM)-ANFIS method</i>	2018	Malásia	Atração do Cliente	(YADEGARIDEHKORDI et al., 2018)
E44	<i>Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies</i>	2012	Paquistão	Retenção do Cliente	(IDRIS; RIZWAN; KHAN, 2012)
E45	<i>Predicting online channel acceptance with social network data</i>	2014	Bélgica	Desenvolvimento do Cliente	(VERBRAKEN et al., 2014)
E46	<i>Late payment prediction models for fair allocation of customer contact lists to call center agents</i>	2014	Coreia do Sul	Identificação do Cliente	(KIM; KANG, 2016)
E47	<i>Improved churn prediction in telecommunication industry using data mining techniques</i>	2014	Irã	Retenção do Cliente	(KERAMATI et al., 2014)
E48	<i>Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach</i>	2013	China	Desenvolvimento do Cliente	(CHEN; FAN, 2013)
E49	<i>An effective hybrid learning system for telecommunication churn prediction</i>	2013	Irlanda	Retenção do Cliente	(HUANG; KECHADI, 2013)
E50	<i>Improved churn prediction in telecommunication industry by analyzing a large network</i>	2014	Coreia do Sul	Retenção do Cliente	(KIM; JUN; LEE, 2014)
E51	<i>Dynamic churn prediction framework with more effective use of rare event data: The case of private banking</i>	2014	Turquia	Identificação do Cliente	(ALI; ARITÜRK, 2014)

Continua

ID	Título	Ano	País	Dimensão	Referência
E52	<i>A hybrid fuzzy-based personalized recommender system for telecom products/services</i>	2013	Austrália	Desenvolvimento do Cliente	(ZHANG et al., 2013)
E53	<i>Preprocessing unbalanced data using support vector machine</i>	2012	Hong Kong	Identificação do Cliente	(FARQUAD; BOSE, 2012)
E54	<i>A new model for assessment fast food customer behavior case study: An Iranian fast-food restaurant</i>	2013	Irã	Identificação do Cliente	(MOMTAZ; ALIZADEH; VAGHEFI, 2013)
E55	<i>Improving inventory performance with clustering based demand forecasts</i>	2012	Índia	Identificação do Cliente	(BALA, 2012)
E56	<i>Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior</i>	2017	Irã	Identificação do Cliente	(KHODABANDEHLOU; RAHMAN, 2017)
E57	<i>Gaining customer knowledge in low cost airlines through text mining</i>	2014	Malásia	Atração do Cliente	(LIAU; TAN, 2014)
E58	<i>Dimensionality and data reduction in telecom churn prediction</i>	2014	Taiwan	Identificação do Cliente	(LIN; TSAI; KE, 2014)
E59	<i>Durable product review mining for customer segmentation</i>	2015	China	Identificação do Cliente	(JIANG et al., 2015)
E60	<i>A new recommendation system on the basis of consumer initiative decision based on associative classification approach</i>	2018	China	Desenvolvimento do Cliente	(YIN et al., 2018)
E61	<i>Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods</i>	2016	Irã	Identificação do Cliente	(FATHIAN; HOSEINPOOR; MINAEI-BIDGOLI, 2016)
E62	<i>Improving customer retention management through cost-sensitive learning</i>	2014	França	Retenção do Cliente	(COUSSEMENT, 2014)
E64	<i>Using customer-related data to enhance e-grocery home delivery</i>	2017	França	Retenção do Cliente	(PAN et al., 2017)
E65	<i>Identifying high-value airlines customers for strategies of online marketing systems: An empirical case in Taiwan</i>	2018	Taiwan	Identificação do Cliente	(CHIANG, 2018b)

Continua

ID	Título	Ano	País	Dimensão	Referência
E66	<i>Customer segmentation for East African Microgrid Consumers</i>	2018	Ruanda	Identificação do Cliente	(OTIENO; WILLIAMS; MCSHARRY, 2018)
E67	<i>Prediction of Churning Behavior of Customers in Telecom Sector Using Supervised Learning Techniques</i>	2018	Paquistão	Retenção do Cliente	(ALI et al., 2018)
E68	<i>Predicting Inactiveness in Telecom (Prepaid) Sector: A Complex Bigdata Application</i>	2018	Índia	Identificação do Cliente	(VYAS et al., 2018)
E69	<i>User Satisfaction Management in E-Government: One K-means Algorithm-based Analysis</i>	2018	China	Retenção do Cliente	(LIU, 2018)
E70	<i>A Systematic Approach to Customer Segmentation and Buyer Targeting for Profit Maximization</i>	2018	Índia	Identificação do Cliente	(BHADE et al., 2018)
E71	<i>Customer support ticket escalation prediction using feature engineering</i>	2018	Canadá	Retenção do Cliente	(MONTGOMERY et al., 2018)
E72	<i>Capturing user sentiments for online Indian movie reviews: A comparative analysis of different machine-learning models</i>	2018	Índia	Identificação do Cliente	(TRIVEDI; DEY; KUMAR, 2018)
E73	<i>Solving customer insurance coverage sales plan problem using a multi-stage data mining approach</i>	2018	Irã	Retenção do Cliente	(ABDI; KHALILI-DAMGHANI; ABOLMAKAREM, 2018)
E74	<i>Applying data mining for online CRM marketing strategy: An empirical case of coffee shop industry in Taiwan</i>	2018	Taiwan	Desenvolvimento do Cliente	(CHIANG, 2018a)
E75	<i>A two-phase multiobjective optimization for benchmarking and evaluating service quality in banks</i>	2018	Nigéria	Retenção do Cliente	(AYO, 2018)
E76	<i>Forecasting the next likely purchase events of insurance customers: A case study on the value of data-rich multi-channel environments</i>	2018	Suíça	Identificação do Cliente	(MAU; PLETIKOSA; WAGNER, 2018)
E77	<i>An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications</i>	2018	China	Retenção do Cliente	(YANG et al., 2018)

Continua

ID	Título	Ano	País	Dimensão	Referência
E78	<i>A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees</i>	2018	França	Retenção do Cliente	(CAIGNY; COUSSEMENT; BOCK, 2018)
E79	<i>Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines</i>	2018	China	Retenção do Cliente	(WANG; LU; TAN, 2018)
E80	<i>Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries</i>	2018	Irã	Identificação do Cliente	(KHALILI-DAMGHANI; ABDI; ABOLMAKAREM, 2018)
E81	<i>Customers Segmentation in the Insurance Company (TIC) Dataset</i>	2018	Emirados Árabes Unidos	Identificação do Cliente	(QADADEH; ABDALLAH, 2018)
E82	<i>Probabilistic baseline estimation based on load patterns for better residential customer rewards</i>	2018	Estados Unidos	Retenção do Cliente	(WENG; YU; RAJAGOPAL, 2018)
E83	<i>Retail business analytics: Customer visit segmentation using market basket data</i>	2018	Grécia	Identificação do Cliente	(GRIVA et al., 2018)
E84	<i>Analyzing high speed rail passengers' train choices based on new online booking data in China</i>	2018	Estados Unidos	Identificação do Cliente	(SUN et al., 2018)

Fonte: Elaborado pela Autora.

B LISTA DE ESTUDOS RETORNADOS NA ETAPA DE SELEÇÃO

ID	Título	Ano	Referências
E1	<i>Neural Network and Classification Approach in Identifying Customer Behavior in the Banking Sector: A Case Study of an International Bank</i>	2015	(OGWUELEKA et al., 2015)
E2	<i>Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis</i>	2016	(SARVARI; USTUNDAG; TAKCI, 2016)
E3	<i>A hybrid customer prediction system based on multiple forward stepwise logistic regression mode</i>	2012	(SOROUSH; BAHREININEJAD; BERG, 2012)
E4	<i>Support vector clustering for customer segmentation on mobile tv service</i>	2015	(ALBUQUERQUE; ALFINITO; TORRES, 2015)
E5	<i>Supervised massive data analysis for telecommunication customer churn prediction</i>	2016	(LI et al., 2016)
E6	<i>Handling class imbalance in customer behavior prediction</i>	2014	(LIU et al., 2014)
E7	<i>A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data</i>	2017	(SEYFIOĞLU; DEMIREZEN, 2017)
E8	<i>A customer churn prediction model in telecom industry using boosting</i>	2014	(LU et al., 2014)
E9	<i>Analysis of customer churn prediction in telecom industry using decision trees and logistic regression</i>	2016	(DALVI et al., 2016)
E10	<i>Resolving class imbalance and feature selection in customer churn dataset</i>	2017	(HANIF; AZHAR, 2017)
E11	<i>Customer churn prediction in an internet service provider</i>	2017	(DO et al., 2017)
E12	<i>Echo state network with SVM-readout for customer churn prediction</i>	2015	(RODAN; FARIS, 2015)
E13	<i>Classifying and understanding prospective customers via heterogeneity of supermarket Stores</i>	2017	(TANAKA et al., 2017)
E14	<i>Targeting high value customers while under resource constraint: partial order constrained optimization with genetic algorithm</i>	2015	(CUI; WONG; WAN, 2015)

Continua

ID	Título	Ano	Referências
E15	<i>ERP Neural Network Inventory Control</i>	2017	(FARHAT; OWAY-JAN, 2017)
E16	<i>Ontology-based reasoning for the intelligent handling of customer complaints</i>	2015	(LEE; WANG; TRAPPEY, 2015)
E17	<i>A decision-making framework for precision marketing</i>	2015	(YOU et al., 2015)
E18	<i>Revenue prediction by mining frequent itemsets with customer analysis</i>	2017	(WENG, 2017)
E19	<i>Customer churn prediction in the telecommunication sector using a rough set approach</i>	2017	(AMIN et al., 2017)
E20	<i>Segmenting customers in online stores based on factors that affect the customer's intention to purchase</i>	2012	(HONG; KIM, 2012)
E21	<i>The added value of social media data in B2B customer acquisition systems: A real-life experiment</i>	2017	(MEIRE; BALLINGS; POEL, 2017)
E22	<i>The BIG CHASE: A decision support system for client acquisition applied to financial networks</i>	2017	(QUIJANO-SANCHEZ; LIBERATORE, 2017)
E23	<i>Using the Taguchi method for effective market segmentation</i>	2012	(HONG, 2012)
E24	<i>Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip</i>	2017	(MORO; RITA; COELHO, 2017)
E25	<i>Integrating expert knowledge and multilingual web crawling data in a lead qualification system</i>	2016	(D'HAEN et al., 2016)
E26	<i>Managing B2B customer churn, retention and profitability</i>	2014	(JAHROMI; STAKHOVYCH; EWING, 2014)
E27	<i>An efficient CRM-data mining framework for the prediction of customer behaviour</i>	2015	(BAHARI; ELAYIDOM, 2015)
E28	<i>A comparison of machine learning techniques for customer churn prediction</i>	2015	(VAFEIADIS et al., 2015)
E29	<i>Early churn prediction with personalized targeting in mobile social games</i>	2017	(MILOŠEVIĆ; ŽIVIĆ; ANDJELKOVIĆ, 2017)
E30	<i>A multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations</i>	2016	(CHEN; FAN; SUN, 2016)

Continua

ID	Título	Ano	Referências
E31	<i>Churn prediction using comprehensible support vector machine: An analytical CRM application</i>	2014	(FARQUAD; RAVI; RAJU, 2014)
E32	<i>An Artificial Neural Network-based technique for on-line hotel booking</i>	2014	(CORAZZA; FASANO; MASON, 2014)
E33	<i>Predictive segmentation of energy consumers</i>	2016	(ALBERT; MAASOUMY, 2016)
E34	<i>The added value of auxiliary data in sentiment analysis of Facebook posts</i>	2016	(MEIRE; BALLINGS; POEL, 2016)
E35	<i>GMDH-based semi-supervised feature selection for customer classification</i>	2017	(XIAO et al., 2017)
E36	<i>Distributed customer behavior prediction using multi-plex data: a collaborative MK-SVM approach</i>	2012	(CHEN; FAN, 2012)
E37	<i>Detecting the migration of mobile service customers using fuzzy clustering</i>	2015	(BOSE; CHEN, 2015)
E38	<i>The classification of internet shop customers based on the cluster analysis and graph cellular automata</i>	2017	(MAŁECKI; WĄTRÓBSKI, 2017)
E39	<i>Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach</i>	2013	(ZHENG; ZHU; LIN, 2013)
E40	<i>Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales</i>	2013	(DUCHESSI; LAURÍA, 2013)
E41	<i>Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis</i>	2017	(FAN; CHE; CHEN, 2017)
E42	<i>On the brink: Predicting business failure with mobile location-based checkins</i>	2015	(WANG et al., 2015)
E43	<i>Predicting determinants of hotel success and development using Structural Equation Modelling (SEM)-ANFIS method</i>	2018	(YADEGARIDEHKORDI et al., 2018)
E44	<i>Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies</i>	2012	(IDRIS; RIZWAN; KHAN, 2012)
E45	<i>Predicting online channel acceptance with social network data</i>	2014	(VERBRAKEN et al., 2014)

Continua

ID	Título	Ano	Referências
E46	<i>Late payment prediction models for fair allocation of customer contact lists to call center agents</i>	2014	(KIM; KANG, 2016)
E47	<i>Improved churn prediction in telecommunication industry using data mining techniques</i>	2014	(KERAMATI et al., 2014)
E48	<i>Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach</i>	2013	(CHEN; FAN, 2013)
E49	<i>An effective hybrid learning system for telecommunication churn prediction</i>	2013	(HUANG; KECHADI, 2013)
E50	<i>Improved churn prediction in telecommunication industry by analyzing a large network</i>	2014	(KIM; JUN; LEE, 2014)
E51	<i>Dynamic churn prediction framework with more effective use of rare event data: The case of private banking</i>	2014	(ALI; ARITÜRK, 2014)
E52	<i>A hybrid fuzzy-based personalized recommender system for telecom products/services</i>	2013	(ZHANG et al., 2013)
E53	<i>Preprocessing unbalanced data using support vector machine</i>	2012	(FARQUAD; BOSE, 2012)
E54	<i>A new model for assessment fast food customer behavior case study: An Iranian fast-food restaurant</i>	2013	(MOMTAZ; ALI-ZADEH; VAGHEFI, 2013)
E55	<i>Improving inventory performance with clustering based demand forecasts</i>	2012	(BALA, 2012)
E56	<i>Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior</i>	2017	(KHODABANDEHLOU; RAHMAN, 2017)
E57	<i>Gaining customer knowledge in low cost airlines through text mining</i>	2014	(LIAU; TAN, 2014)
E58	<i>Dimensionality and data reduction in telecom churn prediction</i>	2014	(LIN; TSAI; KE, 2014)
E59	<i>Durable product review mining for customer segmentation</i>	2015	(JIANG et al., 2015)
E60	<i>A new recommendation system on the basis of consumer initiative decision based on associative classification approach</i>	2018	(YIN et al., 2018)

Continua

ID	Título	Ano	Referências
E61	<i>Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods</i>	2016	(FATHIAN; HOSEIN-POOR; MINAEI-BIDGOLI, 2016)
E62	<i>Improving customer retention management through cost-sensitive learning</i>	2014	(COUSSEMENT, 2014)
E63	<i>Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco</i>	2016	(TSE et al., 2016)
E64	<i>Using customer-related data to enhance e-grocery home delivery</i>	2017	(PAN et al., 2017)
E65	<i>Identifying high-value airlines customers for strategies of online marketing systems: An empirical case in Taiwan</i>	2018	(CHIANG, 2018b)
E66	<i>Customer segmentation for East African Microgrid Consumers</i>	2018	(OTIENO; WILLIAMS; MCSHARRY, 2018)
E67	<i>Prediction of Churning Behavior of Customers in Telecom Sector Using Supervised Learning Techniques</i>	2018	(ALI et al., 2018)
E68	<i>Predicting Inactiveness in Telecom (Prepaid) Sector: A Complex Bigdata Application</i>	2018	(VYAS et al., 2018)
E69	<i>User Satisfaction Management in E-Government: One K-means Algorithm-based Analysis</i>	2018	(LIU, 2018)
E70	<i>A Systematic Approach to Customer Segmentation and Buyer Targeting for Profit Maximization</i>	2018	(BHADE et al., 2018)
E71	<i>Customer support ticket escalation prediction using feature engineering</i>	2018	(MONTGOMERY et al., 2018)
E72	<i>Capturing user sentiments for online Indian movie reviews: A comparative analysis of different machine-learning models</i>	2018	(TRIVEDI; DEY; KUMAR, 2018)
E73	<i>Solving customer insurance coverage sales plan problem using a multi-stage data mining approach</i>	2018	(ABDI; KHALILI-DAMGHANI; ABOLMAKAREM, 2018)
E74	<i>Applying data mining for online CRM marketing strategy: An empirical case of coffee shop industry in Taiwan</i>	2018	(CHIANG, 2018a)
E75	<i>A two-phase multiobjective optimization for benchmarking and evaluating service quality in banks</i>	2018	(AYO, 2018)

Continua

ID	Título	Ano	Referências
E76	<i>Forecasting the next likely purchase events of insurance customers: A case study on the value of data-rich multichannel environments</i>	2018	(MAU; PLETIKOSA; WAGNER, 2018)
E77	<i>An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications</i>	2018	(YANG et al., 2018)
E78	<i>A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees</i>	2018	(CAIGNY; COUSSEMENT; BOCK, 2018)
E79	<i>Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines</i>	2018	(WANG; LU; TAN, 2018)
E80	<i>Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries</i>	2018	(KHALILI-DAMGHANI; ABDI; ABOLMAKAREM, 2018)
E81	<i>Customers Segmentation in the Insurance Company (TIC) Dataset</i>	2018	(QADADEH; ABDALLAH, 2018)
E82	<i>Probabilistic baseline estimation based on load patterns for better residential customer rewards</i>	2018	(WENG; YU; RAJAGOPAL, 2018)
E83	<i>Retail business analytics: Customer visit segmentation using market basket data</i>	2018	(GRIVA et al., 2018)
E84	<i>Analyzing high speed rail passengers' train choices based on new online booking data in China</i>	2018	(SUN et al., 2018)

Fonte: Elaborado pelo Autora.

C DETALHES DA AVALIAÇÃO DE QUALIDADE

Ref.	P1	P2	P3	P4	Total
E1	S	S	S	S	4.0
E2	S	S	S	S	4.0
E3	S	S	S	S	4.0
E4	S	S	P	S	3.5
E5	S	S	P	S	3.5
E6	S	S	S	S	4.0
E7	S	S	S	S	4.0
E8	S	S	S	S	4.0
E9	S	S	P	P	3.0
E10	S	P	S	S	3.5
E11	S	P	S	S	3.5
E12	S	S	S	S	4.0
E13	S	S	S	S	4.0
E14	S	S	S	P	3.5
E15	S	P	P	S	3.0
E16	S	S	S	S	4.0
E17	S	S	S	S	4.0
E18	S	S	S	S	4.0
E19	S	S	S	S	4.0
E20	S	S	S	S	4.0
E21	S	S	S	S	4.0
E22	S	S	S	S	4.0
E23	S	S	S	S	4.0
E24	P	S	S	P	3.0
E25	S	P	S	S	3.5
E26	S	S	S	S	4.0
E27	S	P	S	S	3.5
E28	S	S	S	S	4.0
E29	S	S	S	S	4.0
E30	S	S	S	S	4.0
E31	S	S	S	S	4.0
E32	S	S	P	S	3.5
E33	S	S	S	S	4.0
E34	S	S	S	S	4.0
E35	S	S	S	S	4.0

Continua

Ref.	P1	P2	P3	P4	Total
E36	S	S	S	S	4.0
E37	S	S	S	S	4.0
E38	S	S	P	P	3.0
E39	S	S	S	S	4.0
E40	S	S	P	S	3.5
E41	S	S	S	S	4.0
E42	S	S	S	S	4.0
E43	S	S	S	S	4.0
E44	S	S	S	S	4.0
E45	S	S	S	S	4.0
E46	S	S	S	S	4.0
E47	S	S	S	S	4.0
E48	S	S	S	S	4.0
E49	S	S	S	S	4.0
E50	S	S	S	S	4.0
E51	S	S	S	S	4.0
E52	S	S	S	S	4.0
E53	S	S	S	S	4.0
E54	S	S	P	S	3.5
E55	S	S	S	S	4.0
E56	S	S	S	S	4.0
E57	S	S	P	S	3.5
E58	S	S	S	S	4.0
E59	S	S	P	S	3.5
E60	S	S	S	S	4.0
E61	S	S	S	S	4.0
E62	S	S	S	S	4.0
E63	N	N	N	N	0.0
E64	P	S	S	P	3.0
E65	S	S	S	S	4.0
E66	S	S	S	S	4.0
E67	S	S	S	S	4.0
E68	S	S	S	S	4.0
E69	S	S	S	S	4.0
E70	S	S	S	S	4.0
E71	S	S	S	S	4.0

Continua

Ref.	P1	P2	P3	P4	Total
E72	S	S	S	S	4.0
E73	S	S	S	S	4.0
E74	S	S	P	S	3.0
E75	S	S	P	S	3.5
E76	S	S	S	S	4.0
E77	S	S	S	S	4.0
E78	S	S	S	S	4.0
E79	S	S	S	S	4.0
E80	S	S	S	S	4.0
E81	S	S	P	S	3.5
E82	S	S	S	S	4.0
E83	S	S	S	S	4.0
E84	S	S	P	S	3.5

Fonte: Elaborado pela Autora.

Current Applications of Machine Learning Techniques in CRM: *A Literature Review and Practical Implications*

Beatriz Nery Rodrigues Chagas
State University of Maranhão
São Luís, Brazil
beatriznery.12@gmail.com

Julio Viana
University of Leipzig
Leipzig, Germany
julio.viana@scrc-leipzig.de

Olaf Reinhold
University of Leipzig
Leipzig, Germany
reinhold@wifa.uni-leipzig.de

Fabio Lobato
Federal University of Western of Pará
Santarém, Brazil
fabio.lobatof@gmail.com

Antonio F. L. Jacob Jr.
State University of Maranhão
São Luís, Brazil
antonio.jacob@gmail.com

Rainer Alt
University of Leipzig
Leipzig, Germany
rainer.alt@uni-leipzig.de

Abstract—Customer Relationship Management (CRM) has been an important ally to companies willing to improve their customer experiences. Lately, Machine Learning (ML) and its techniques have been transforming the way companies interact with their customers through data analysis. This paper reviews the literature on the application of ML techniques to improve CRM processes and provides an overview of used techniques and their application to each CRM dimension and element. In addition, the practical implications are analyzed based on recent developments of CRM tools in the field of ML. The paper delivers insights for future studies as well as for CRM tools willing to improve their analytical and automated services and features.

CRM, Machine Learning, Information Systems.

I. INTRODUCTION

Companies are constantly seeking to differentiate their brands, products and services. Creating a superior customer experience along the entire value chain in a cost-efficient way is one of the main challenges for businesses nowadays [1] [2].

When designing data-driven strategies to create a unique customer experience, companies rely on the concepts of the well-established field of Customer Relationship Management (CRM). CRM "utilizes technology as an enabler to capture, analyze and disseminate current and prospective customer data to identify customer needs more precisely and to develop insightful relationships" [3]. This implies the integration of strategic goals with business processes and Information and Communication Technology (ICT) [4]. CRM has its basis on the need for extensive knowledge about customers and the ability to enrich and use it in operative and analytical processes.

Concepts, such as Artificial Intelligence (AI), Computational Intelligence (CI), Machine Learning (ML) and Data Mining (DM) brought to the market algorithms and techniques to process and use data. AI is a multidisciplinary science that seeks to develop and apply computational techniques that simulate human behavior [5]. CI, however, aims at using learning, adaptive or evolutionary computing to create software that is, in some sense, intelligent [6]. CI considers the implementation and development of neural networks, evolutionary algorithms, fuzzy-logic-based systems and hybrid intelligent systems [7], [8]. Both complementary fields use the methodology of ML to classify and make predictions based on a certain database.

For this study, we consider ML as a methodology that gives computers the ability to learn without explicit programming. It relies on the premise that machines should learn with

the data samples. Basically, ML consists of three main types of learning: supervised, unsupervised and semi-supervised [9], [10], [11], [12], [13].

Studies announce ML as a key issue for future businesses [14] [15]. However, while some studies investigated the role of DM in CRM, the literature lacks information about the state-of-art research about the application fields and benefits of ML for CRM.

This paper analyzes the current literature to identify techniques and applications of ML applied to CRM processes and discusses the practical implications of these applications to CRM tools. It aims at identifying research gaps and improving existing tools based on the current landscape.

This paper seeks to answer the following research questions:

- a) What are the ML techniques applied to CRM?
- b) How do ML techniques improve CRM processes (Practical Implications)?

For that, we analyzed the literature systematically to extract current techniques and applications, as well as an overview of these applications in CRM tools. At first, the article introduces the research methodology and presents an overview about the findings from literature. Subsequently, existing CRM tools are reviewed regarding the application fields of the identified techniques in order to discuss the practical implications and contributions of ML to the field of CRM.

II. SYSTEMATIC LITERATURE REVIEW

A systematic literature review is a formal process adopted to identify, evaluate, and synthesize evidences of empirical results to answer a research question. Performing a systematic review involves several discrete activities, which can be grouped into three main phases: planning, conducting the review, and the reporting the review [16].

Besides answering research questions, a protocol is defined. This protocol consists of a research agenda that describes the process while planning the systematic review. It is a documentation of the review process and must have all the information needed to perform it in a systematic way [16], [17]. The research agenda includes all steps to reach to the desirable results. Figure 1 depicts this agenda, introducing the search steps, the selection and analysis of the literature. It also synthesizes the number of accepted articles.

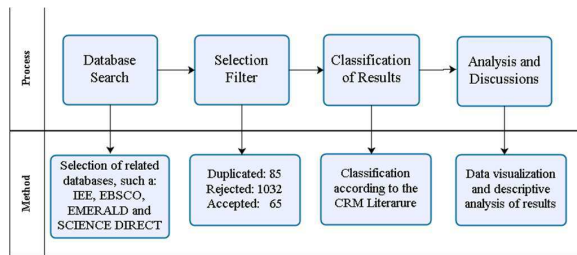


Fig. 1. Research Agenda.

A. Conceptualization of the Study

This study provides an overview of current ML techniques and their application to the field of CRM. Hence, the main areas of research goes around the concepts of ML and CRM.

It is common to find the term ML together with CI and AI, since its techniques are often related to the integration of two or more DM algorithms [12]. Therefore, a clear definition for these terms is essential to the process of literature review. Table I introduces the terms and serves as a guide to the research process and selection of relevant research articles.

TABLE I. Definitions of Terms Related to the Research

Item	Definition	References
CRM	CRM is a strategic approach that aims at creating improved shareholder value. CRM unites the potential of relationship marketing strategies and ICT to create profitable, long-term relationships with customers.	[4]
AI	AI is a multidisciplinary science, in which computational methods are developed to simulate human behavior in specific activities.	[5]
CI	CI, as a branch of the study of AI, aims to use learning and optimization methods to develop intelligent systems.	[6]
ML	ML is a methodology that gives computers the ability to learn without explicit programming. ML is included in the fields of AI or CI, depending on the used algorithms or techniques. It relies on the premise that machines should learn with the provided data, as of their experiences. ML uses different techniques from the field of DM, including some extra learning algorithms.	[9], [10], [11], [12], [13]
DM	DM focuses on the extraction of knowledge and insights through the analysis of large amounts of data, using sophisticated modeling techniques. It converts data into knowledge and actionable information. A DM procedure has many stages. It typically involves extensive data management before the application of a statistical or ML algorithm and the development of an appropriate model.	[18]

Research concerning the topics of ML and CRM can be scattered in various journals across different disciplines, such as Information Technology, Information Systems, Marketing, Business Administration, etc. Therefore, this research presents results from different databases in order to integrate the most relevant research on the topic. The research took place using the following databases:

- Ebsco Database
- IEEE Xplore
- Science Direct
- Emerald

These four databases provided a comprehensive bibliography of academic literature across the relevant disciplines in the topics of ML and CRM.

Since ML is commonly used within the topic of AI or CI, a broad search was performed to identify possible missing applications. The addition of the term ‘Artificial Intelligence’ was important to certify that the use of different terminologies would not interfere negatively in the results.

Based on the definitions in Table I, the following search phrases were applied in the aforementioned databases:

- Machine Learning + Customer Relationship Management + Algorithm
- Machine Learning + Customer Relationship Management + Technique
- Artificial Intelligence + Customer Relationship Management + Algorithm
- Artificial Intelligence + Customer Relationship Management + Technique

The terms ‘algorithm’ and ‘techniques’ were used in order to filter papers that contain the application of an algorithm or technique from ML to solve an organizational issue regarding CRM.

B. Classification and Filtering

The first search resulted in 1182 hits. These hits were filtered according to the application of the algorithms or techniques to the field of CRM. For this purpose, we considered the four dimensions and their elements in CRM introduced in Table II. The strategic goals of CRM are executed through CRM processes for marketing, sales and service purposes [19]. The process of building up a customer relationship can be structured in the four dimensions of customer identification, customer attraction, customer retention and customer development. They are based on an accurate and deep understanding of the customer [20].

TABLE II. The Four Dimensions of CRM.

Dimensions	Elements
Customer Identification	Target Customer Analysis Customer Segmentation
Customer Attraction	Direct Marketing
Customer Retention	Loyalty Program One-to-One Marketing Complaint Management
Customer Development	Customer Lifetime Cycle Up/Cross Selling Market Basket Analysis

Customer Identification: consists on the first step of the CRM framework. It seeks to target possible customers by identifying most profitable ones and involves their classification or segmentation [21].

Customer Attraction: allocates resources to attract the identified target groups. Customer attraction involves the use of an appropriate method of communication and the elimination of any sort of wasted effort [22], [23].

Customer Retention: intends to realize the customer expectations satisfaction with customer perceptions of satisfaction is the basic condition for maintaining the customer. Elements of this dimension include one-to-one marketing, loyalty and bonus programs, and complaints management [23], [24].

Customer Development: focuses on increasing transaction intensity, transaction value, and individual customer profitability. The main elements of this dimension are customer lifetime, value analysis and up/cross selling [25], [23].

Additionally, we considered the three phases of ML (training, validation and testing) to filter the results and select papers that describe an application of ML to CRM. When applying these techniques, researchers and practitioners perform the following four steps:

- Identify the features from training data;
- Identify subset of attributes necessary for classification (e.g.: dimensional reduction);
- Learn model using training data;
- Use trained model to classify unknown data and predict the result accurately [26].

Using these steps to select the appropriate application of ML techniques, we filtered research papers considered relevant to the analysis. Besides using the aforementioned definitions to identify and filter papers of interest, we applied further filters as follows:

- Time of publication restricted to articles published within the last 6 years, following the example of other literature reviews in the field of AI and ML [27], [28].
- Exclusion of repeated articles.
- Exclusion of conference papers, master and doctoral thesis, textbooks, websites and unfinished working papers.

C. Discussions of Results

Considering the filters applied, we reached to a final number of unique and relevant 'hits'.

TABLE III. Total Number of Relevant and Unique Articles.

Database	Search Phrase A	Search Phrase B	Search Phrase C	Search Phrase D	Unique relevant hits
Ebsco	5 (0)	4 (1)	3 (2)	6 (1)	4
IEEE	12 (1)	7 (0)	17 (4)	10 (3)	8
S. Direct	287 (18)	293 (8)	202 (14)	233 (1)	41
Emerald	25 (8)	24 (2)	22 (1)	32 (1)	12
TOTAL					65

Table III presents the number of papers for each database and search phrases. The number in brackets (hits) represents the selected articles after manual evaluation and application of filters. The numbers dropped significantly after the analysis and application of defined filters, especially when considering the application of ML techniques to CRM dimensions. The process resulted in 65 net articles.

From this selection, a classification by ML techniques was carried out in relation to the CRM elements. Table IV introduces the techniques and the corresponding papers.

TABLE IV. Classification of the ML Techniques in CRM Elements.

CRM Elements	Techniques	References
Target Customer Analysis	Neural Network; Multiple Logistic Regression; Weighted Random Forest; RUSBoost; Genetic Algorithm; Decision Tree; SVM (Support Vector Machine).	[29], [30], [31], [32], [33], [34], [35], [36], [37].
Customer Segmentation	Self-organizing Maps; Support Vector Clustering; XG-Boost; Logistic Regression; Naïve Bayes; Neural Network; C-MK-SVM (Collaborative Multiple Kernel Support Vector Machine); Random Forest; SVM; PCA (Principal Component Analysis).	[38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54].

Direct Marketing	Logistic Regression; Projected Gradient Descent Algorithm; Fuzzy.	[55], [56], [57], [58], [59].
Loyalty Program	Random Forest; Decision Tree; Logistic Regression; Fuzzy.	[60], [61], [62], [63].
One-to-one Marketing	SVM; Exhaustive Algorithm, Genetic Algorithm, Covering Algorithm; Decision Tree, Logistic Regression; Gradient Boosting; Naïve Bayes; Random Forest; PCA; Neural Network.	[64], [65], [66], [67], [68], [69], [70], [71], [72].
Complaint Management	Logistic Regression; Random Forest, Gradient Boosting; XG-Boost, AdaBoost, Neural Networks, Extra Trees; Case-based Reasoning; SVM; Naïve Bayes; Decision Tree; NBTree (Naïve Bayes Tree); Co-EM Ensemble; Co-EM SVM; Co-training; Co-EM Bayesian.	[73], [74], [75], [76], [77], [78], [79], [80], [81], [82].
Customer Lifetime Cycle	Random Forest; SVM; Naïve Bayes; MK-SVR.	[83], [57], [84], [85], [47].
Up/Cross Selling	MK-STM (Multi-kernel Support Tensor Machine); Neural Network; Fuzzy Algorithm.	[86], [87], [88], [89].
Market Basket Analysis	Neural Network; Logit Model; K-Nearest Neighbor.	[90], [91].

Customer Identification and Retention are the most discussed CRM dimensions in the literature, as the number of articles describing the application of techniques and algorithms in these areas is higher than the other two dimensions. Regarding the CRM elements, Customer Segmentation has the highest number of papers using ML techniques to improve the clustering process. Table V classifies the results according to the dimensions and elements.

TABLE V. Number of Articles according to CRM Dimensions and Elements.

Dimensions	Elements	Articles	Articles
Customer Identification	Target Customer Analysis	26	9
	Customer Segmentation		17
Customer Attraction	Direct Marketing	5	5
Customer Retention	Loyalty Program	22	4
	One-to-One Marketing		8
	Complaint Management		10
Customer Development	Customer Lifetime Cycle	11	5
	Up/Cross Selling		4
	Market Basket Analysis		2

The selected articles considered the use of ML techniques and their phases to solve CRM problems. For that, many of these works applied well-known techniques, such as Naïve Bayes, SVM, Artificial Neural Networks, K-Means, Logistic Regression and others. However, some papers presented new algorithms and compared them with the existing ones. The next table shows the techniques and the number of articles using these approaches. The high number of 25 other techniques consists mostly of combinations of different algorithms to solve particular problems.

TABLE VI. Most Used ML Algorithms in Accepted Papers.

Techniques	Number of Papers
SVM	13
Neural Network	12
Logistic Regression	12
Decision Tree	10
K-Means	9
Random Forest	8

K-Nearest Neighbor	7
Naïve Bayes	5
Self-Organizing Maps	4
PCA	4
AdaBoost	3
Genetic Algorithm	2
XG-Boost	2
Gradient Boosting	2
Others	25

This study considered articles produced in the last six years. The number of research regarding ML application in CRM has been increasing in the past years, mirroring the increase of discussions in the industry and academia. Fig. 2 presents this analysis and shows the peak in 2017, with 18 selected papers. The low number in 2018 reflects the research period, occurred between March and April 2018.

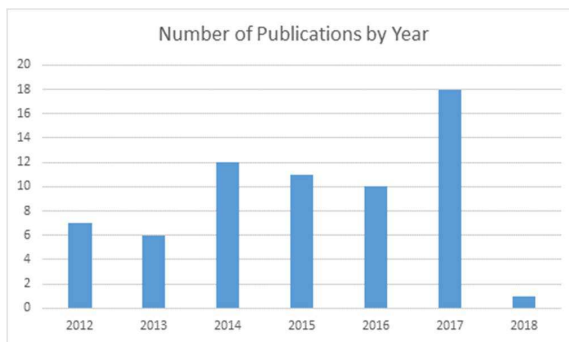


Fig. 2. Number of Articles by Year of Publication.

III. PRACTICAL IMPLICATIONS

The analysis revealed an increasing use of ML in research. However, most of the papers either present individual solutions for specific problems or consist of academic examples not easily applicable by other businesses in general.

Nevertheless, we selected the examples of papers that explicitly stated the practical implications of their research. Table VII presents the practical implications related to the focused CRM dimensions.

TABLE VII. Practical Implications from the Literature and Corresponding CRM Dimensions.

Ref.	Practical Implications	Dimension
[42]	Online stores can develop promotional marketing and offer personalized service for e-customers.	Customer Attraction
[59]	Knowledge gain of customer sentiments and improvements on customer service to attract more customers and generate more profits.	Customer Attraction
[83]	Precise supply strategies for different customer categories to target high potential customers.	Customer Development
[90]	Simulation of a system that predicts good demand for an enterprise's inventory according to sales.	Customer Development
[30]	Technique considerably reduces computational costs, providing more accurate prediction.	Customer Identification
[35]	Customer behavior analysis to understand needs and establish strong relationships (fast food sector).	Customer Identification
[38]	Marketing and services processes are more effective, by improving performance of segmentation.	Customer Identification
[50]	A model to forecast demand of various items to improve inventory performance.	Customer Identification

[53]	Result can support new product design and development, product repositioning, marketing strategy and product differentiation.	Customer Identification
[63]	Insights to mobile services providers about how to detect temporal changes in customer behavior.	Customer Identification
[29]	Change in Customer behavior patterns to improve customer efficiency, satisfaction and speedy sales.	Customer Retention
[52]	Improving accuracy for churn prediction in the telecom industry.	Customer Retention
[72]	The research emphasizes the importance of cost-sensitive learning to improve customer retention management in the retail context.	Customer Retention
[76]	Complete customer complaint ontology can be applied to customer-oriented service industries such as hospitality, catering, retail, and tourism.	Customer Retention
[77]	Direction of efforts to answer online reviews in alignment with strategies (hotel industry), by profiling the reviews according to member and date.	Customer Retention

The vast majority of papers seeks to combine or improve ML methods and techniques to reach better results compared to previous models, focusing on the contribution of their new developed models. Regardless the short number of 15 papers that clearly stated the practical implications of their research, the smaller sample confirms the results from Table V, with a higher number of papers concerning Customer Identification and Retention.

Despite the practical relevance from Table VII, the number of analyzed articles decreased significantly. For this reason, we investigated current CRM systems and their implementation of ML algorithms. GetCRM analyzes CRM tools and works as a marketplace for these tools with over 50 thousand business users. Their recent study provides insights on the use of AI and ML techniques to improve CRM processes and indicates eight tools that implemented ML [92].

Table VIII shows the tools and the CRM processes improved by the application of ML techniques. The table also indicates the CRM dimensions according to the applications.

TABLE VIII. The Contribution of ML to CRM Dimensions and Tools

CRM Tools	Use of ML Techniques	Addressed CRM Dimensions
Marketo	- Shortens the time for lead qualification.	Customer Identification
Base	- Analyzes data and offers live inside feed, recommendations to increase sales, goal tracking and forecasting.	Customer Attraction and Development
Sugar CRM	- Updates customer information automatically based on external data points. - Recommends specific interactions with customers.	Customer Identification and Development
Zoho	- Detects anomalies in sales operations to predict whether teams will reach their goals. - Predicts optimal time for salespeople to reach out to prospects.	Customer Attraction
Zendesk	- Predicts satisfaction by analyzing the likelihood of particular tickets being reviewed as good or bad.	Customer Retention
Gorgias.io	- Helpdesk to integrate the communication with customers across different channels. - Build workflows that prioritize customers to specific issues.	Customer Retention

Dynamics 365	- Provide translation, sentiment analysis and keyword identification benefiting sales outreach and social customer service	Customer Attraction and Retention
Salesforce	- Learns from data to predict outcomes, prescribe solutions and automate tasks. - Integrates data to generate predictive lead scoring. - Create marketing campaigns.	Customer Identification

As seen, ML provided means for improving processes and analysis in CRM. ML has been supporting these tools in improving process automation, predictions, sales recommendations and facilitating the communication between companies and clients.

In spite of a higher number of studies concerning Customer Identification and Retention indicated in the literature review, the selected CRM tools show a well-distributed focus of ML techniques across the four CRM dimensions. It shows a concern of CRM tools to improve processes in all dimensions and highlights the research gap regarding Customer Attraction and Development.

The advances in academic research can positively influence short-term developments in existing tools, as well as increase the number of new tools. This improvement in CRM processes supports different businesses in refining the relationship with their customers and prospects, influencing the performances of several companies across the globe.

IV. CONCLUSION

This study provides a review on research using ML techniques to solve and/or automate tasks for CRM. Moreover, it provides an overview of the recent developments in current CRM tools. This analysis delivers a better understanding of the practical implications when using ML to improve analysis and processes in CRM.

Scholars across the globe have been discussing the application of ML, its techniques and algorithms in the field of CRM. Additionally, CRM tools are constantly developing new ways of analyzing customer data using ML and AI. The analysis shows that SVM, Neural Network, Logistic Regression and Decision Tree are the most used techniques in ML to tackle CRM issues and improve processes. Additionally, Customer Identification and Retention are the CRM dimensions with the highest concern in the literature.

Besides the literature review, the study provides an analysis of the practical implications and relates these practices to the CRM dimensions. These results are important to understand how ML is able to improve the processes and which tasks in CRM current literature has been targeting. Additionally, it provides insights for companies using ML to develop CRM solutions. This overview can support researchers and practitioners to understand how ML can practically improve CRM activities, as well as the less-focused CRM dimensions.

V. LIMITATIONS AND FURTHER RESEARCH

This study provides an overview of the current applications of ML techniques in CRM processes. However, the literature only pointed out scarcely the benefits of these techniques to CRM processes. Therefore, the analysis of practical implications was important to point out these benefits. A deeper analysis could validate these benefits with experts in

the fields of CRM and ICT and provide better explanation on how ML techniques are incorporated to CRM systems.

The low number of research in some techniques and CRM dimensions hinders a deeper analysis. However, it indicates research gaps for future research.

Recent developments in research and industry also points out to a new field named "Social CRM", which aims at integrating social media data to CRM processes [93]. This integration can benefit from the advances in the field of machine learning, as well as provide new topics for future.

VI. ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of this research by the German Academic Exchange Service (DAAD) with the project *German-Brazilian Partnerships for Social CRM* and the SAB and EU with the EFRE project *Content Provider Shield* (COPS) and ESF project S2DES.

REFERENCES

- [1] P. C. Verhoef and K. N. Lemon, "Understanding Customer Experience Throughout the Customer Journey," *Journal of Marketing: AMA/MSI Special Issue*, vol. 80, pp. 69-96, 2016.
- [2] The Deloitte Consumer Review, "CX Marks the Spot: Rethinking the Customer Experience to Win," Deloitte, London, 2016.
- [3] C. Bull, "Strategic issues in customer relationship management (CRM) implementation," *Business Process Management Journal*, vol. 9, no. 5, pp. 592-602, 2003.
- [4] A. Payne and P. Frow, "A Strategic Framework for Customer Relationship Management," *Journal of Marketing*, vol. 69, no. 4, pp. 167-176, October 2005.
- [5] S. J. Russell, P. Norvig and J. Canny, *Artificial Intelligence: A Modern Approach*, Pearson Education Limited, 2003.
- [6] R. Begg, *Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques*, IGI, 2006.
- [7] Y. S. Ong, M. H. Lim and X. Chen, "Memetic computation—past, present & future [research frontier].," *IEEE Computational Intelligence Magazine*, vol. 5, no. 2, pp. 24-31, 2010.
- [8] A. R. Cárdenas Maita, *Um estudo da aplicação de técnicas de inteligência computacional e de aprendizado em máquina de mineração de processos de negócio*, São Paulo, 2016.
- [9] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1-3, pp. 1-6., 1998.
- [10] L. De Lathauwer, B. De Moor, J. Vandewalle and B. S. S. Higher-Order, "Singular value decomposition," *Proc. EUSIPCO-94*, vol. 1, pp. 175-178, September 1994.
- [11] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, vol. 3, p. 210, 1959.
- [12] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153-1176., 2016.
- [13] K. Mishra and R. G. Rani, *A Machine Learning Approach for Churn Prediction in Telecommunication (Doctoral dissertation)*, 2017.
- [14] C. J. Idoine, P. Krensky, E. Brethenoux, J. Hare, S. Sicular and S. Vashisth, "Magic Quadrant for Data Science and Machine-Learning Platforms," Gartner, Inc, 2018.
- [15] D. Court, "Getting big impact from big data," *McKinsey Quarterly*, 2015.
- [16] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of systems and software*, vol. 4, pp. 571-583, 2007.

- [17] R. A. Sinoara, J. Antunes and S. O. Rezende, "Text mining and semantics: a systematic mapping study," *Journal of the Brazilian Computer Society*, vol. 1, p. 9, 2017.
- [18] K. K. Tsipitsis, and A. Chorianopoulos, Data mining techniques in CRM: inside customer segmentation., John Wiley & Sons., 2011.
- [19] D. Peppers and M. Rogers, Managing Customer Relationships: A Strategic Framework, Hoboken: Wiley, 2004.
- [20] R. S. Swift, Accelerating customer relationships: using CRM and relationship technologies, Upper Saddle River, NJ: Prentice Hall PTR, 2001.
- [21] P. Deepa, "Data mining in CRM-a Survey," *Journal of Chemical and Pharmaceutical Sciences ISSN*, vol. 974, p. 2115.
- [22] A. H. Kracklauer, D. Q. Mills and D. Seifert, "Customer management as the origin of collaborative customer relationship management.," *In Collaborative Customer Relationship Management*, pp. 3-6, 2004.
- [23] V. L. M. Oliveira, "Analytical customer relationship management in retailing supported by data mining techniques," (Doctoral dissertation, Universidade do Porto (Portugal), 2012.
- [24] A. Parvatiyar and J. N. Sheth, "Customer relationship management: Emerging practice, process, and discipline.," *Journal of Economic & Social Research*, vol. 3, no. 2, 2001.
- [25] J. H. Drew, D. R. Mani, A. L. Betz and P. Datta, "Targeting customers with statistical and data-mining techniques.," *Journal of Service Research*, vol. 3, no. 3, pp. 205-219, 2001.
- [26] R. D. Shanbhogue and B. M. Beena, "A survey of data mining and machine learning methods for cyber security intrusion detection.," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153-1176., 2016.
- [27] J. Chai, J. N. Liu and E. W. Ngai, "Application of decision-making techniques in supplier selection: A systematic.," *Expert Systems with Applications*, vol. 40, no. 10, pp. 3872-3885, 2013.
- [28] L. Terán, A. O. Mensah and A. Estorelli, "A literature review for recommender systems techniques used in microblogs," *Expert Systems With Applications*, vol. 103, pp. 63-73, 2018.
- [29] F. Ogwueleka, S. Misra, R. Colomo-Palacios and L. Fernandez, "Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank.," *Human factors and ergonomics in manufacturing & service industries*, vol. 25, no. 1, pp. 28-42, 2015.
- [30] A. Soroush, A. Bahreininejad and J. van den Berg, "A hybrid customer prediction system based on multiple forward stepwise logistic regression mode.," *Intelligent Data Analysis*, vol. 16, no. 2, pp. 265-278, 2012.
- [31] N. Liu, W. Woon, Z. Aung and A. Afshari, "Handling class imbalance in customer behavior prediction," in *In Collaboration Technologies and Systems (CTS), 2014 International Conference on*, 2014.
- [32] G. Cui, M. Wong and X. Wan, "Targeting high value customers while under resource constraint: partial order constrained optimization with genetic algorithm.," *Journal of Interactive Marketing*, vol. 29, pp. 27-37, 2015.
- [33] J. Kim and P. Kang, "Late payment prediction models for fair allocation of customer contact lists to call center agents," *Decision Support Systems*, vol. 85, pp. 84-101, 2016.
- [34] Ö. G. Ali and U. Arıttürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking.," *Expert Systems with Applications*, vol. 41, no. 17, pp. 7889-7903, 2014.
- [35] N. Jafari Momtaz, S. Alizadeh and M. Sharif Vaghefi, , "A new model for assessment fast food customer behavior case study: An Iranian fast-food restaurant.," *British food journal*, vol. 115, no. 4, pp. 601-613, 2013.
- [36] M. Fathian, Y. Hoseinpoor and B. Minaei-Bidgoli, "Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods," *Kybernetes*, vol. 5, pp. 732-743, 2016.
- [37] W. Y. Chiang, "Identifying high-value airlines customers for strategies of online marketing systems: An empirical case in Taiwan," *Kybernetes*, vol. 3, pp. 525-538, 2018.
- [38] P. A. Sarvari, A. Ustundag and H. Takci, "Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis.," *Kybernetes*, vol. 45, no. 7, pp. 1129-1157, 2016.
- [39] P. Albuquerque, S. Alfinito and C. V. Torres, "Support vector clustering for customer segmentation on mobile tv service.," *Communications in Statistics-Simulation and Computation*, vol. 44, no. 6, pp. 1453-1464, 2015.
- [40] M. S. Seyfioglu and M. U. Demirezen, "A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data.," *In Computer Science and Information Systems (FedCSIS)*, pp. 361-365, September 2017.
- [41] T. Tanaka, T. Hamaguchi, T. Saigo and K. Tsuda, "Classifying and Understanding Prospective Customers via Heterogeneity of Supermarket Stores," *Procedia Computer Science*, vol. 112, pp. 956-964., 2017.
- [42] T. Hong and E. Kim, "Segmenting customers in online stores based on factors that affect the customer's intention to purchase.," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2127-2131, 2012.
- [43] C. W. Hong, "Using the Taguchi method for effective market segmentation," *Expert systems with applications*, vol. 39, no. 5, pp. 5451-5459, 2012.
- [44] T. F. Bahari, and M. S. Elayidom., "An efficient CRM-data mining framework for the prediction of customer behaviour.," *Procedia computer science*, vol. 46, pp. 725-731, 2015.
- [45] A. Albert and M. Maasoumy, "Predictive segmentation of energy consumers.," *Applied energy*, vol. 117, pp. 435-448, 2016.
- [46] J. Xiao, H. Cao, X. Jiang, X. Gu and L. Xie, "GMDH-based semi-supervised feature selection for customer classification," *Knowledge-Based Systems*, vol. 132, pp. 236-248, 2017.
- [47] Z. Y. Chen and Z. P. Fan, "Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM approach.," *Knowledge-Based Systems*, vol. 35, pp. 111-119, 2012.
- [48] K. Malecki and J. Wątróbski, "The Classification of Internet Shop Customers based on the Cluster Analysis and Graph Cellular Automata," *Procedia Computer Science*, vol. 112, pp. 2280-2289, 2017.
- [49] M. A. H. Farquad and I. Bose, "Preprocessing unbalanced data using support vector machine," *Decision Support Systems*, vol. 53, no. 1, pp. 226-233, 2012.
- [50] P. K. Bala, "Improving inventory performance with clustering based demand forecasts.," *Journal of Modelling in Management*, vol. 7, no. 1, pp. 23-37, 2012.
- [51] S. Khodabandehlou and M. Zivari Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 1/2, pp. 65-93, 2017.
- [52] W. C. Lin, C. F. Tsai and S. W. Ke, "Dimensionality and data reduction in telecom churn prediction," *Kybernetes*, vol. 43, no. 5, pp. 737-749, 2014.
- [53] S. Jiang, S. Cai, G. Olle Olle and Z. Qin, "Durable product review mining for customer segmentation.," *Kybernetes*, vol. 44, no. 1, pp. 124-138, 2015.
- [54] Y. K. Tse, M. Zhang, B. Doherty, P. Chappell and P. Garnett, "Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco," *Industrial Management & Data Systems*, vol. 6, pp. 1178-1200, 2016.
- [55] J. D'Haen, D. Van den Poel, D. Thorleuchter and D. F. Benoit, "Integrating expert knowledge and multilingual web crawling data in a lead qualification system," *Decision Support Systems*, vol. 82, pp. 69-78, 2016.
- [56] L. Quijano-Sanchez and F. Liberatore, "The BIG CHASE: A decision support system for client acquisition applied to financial networks.," *Decision Support Systems*, vol. 98, pp. 49-58, 2017.
- [57] M. Meire, M. Ballings and D. Van den Poel, "The added value of social media data in B2B customer acquisition systems: A real-life experiment.," *Decision Support Systems*, vol. 104, pp. 26-37, 2017.
- [58] E. Yadegaridehkordi, M. Nilashi, M. H. N. B. M. Nasir and O. Ibrahim, "Predicting determinants of hotel success and development

- using Structural Equation Modelling (SEM)-ANFIS method," *Tourism Management*, vol. 66, pp. 364-386, 2018.
- [59] B. Yee Liao and P. Pei Tan, "Gaining customer knowledge in low cost airlines through text mining," *Industrial Management & Data Systems*, vol. 114, no. 9, pp. 1344-1359, 2014.
- [60] H. Li, D. Yang, L. Yang and Lin, X., "Supervised Massive Data Analysis for Telecommunication Customer Churn Prediction," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, 2016.
- [61] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," in *Colossal Data Analysis and Networking (CDAN), Symposium*, pp. 1-4, March 2016.
- [62] C. H. Weng, "Revenue prediction by mining frequent itemsets with customer analysis," *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 85-97, 2017.
- [63] I. Bose and X. Chen, "Detecting the migration of mobile service customers using fuzzy clustering," *Information & Management*, vol. 52, no. 2, pp. 227-238, 2015.
- [64] A. Rodan and H. Faris, "Echo state network with SVM-readout for customer churn prediction," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on*, pp. 1-5, November 2015.
- [65] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain and K. Huang, "Customer churn prediction in the telecommunication sector using a rough set approach," *Neurocomputing*, vol. 237, pp. 242-254, 2017.
- [66] A. T. Jahromi, S. Stakhovych and M. Ewing, "Managing B2B customer churn, retention and profitability," *Industrial Marketing Management*, vol. 43, no. 7, pp. 1258-1268, 2014.
- [67] M. Milošević, N. Živić and I. Andjelković, "Early churn prediction with personalized targeting in mobile social games," *Expert Systems with Applications*, vol. 83, pp. 326-332, 2017.
- [68] P. Duchessi and E. J. Lauria, "Decision tree models for profiling ski resorts' promotional and advertising strategies and the impact on sales," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5822-5829, 2013.
- [69] A. Idris, M. Rizwan and A. Khan, "Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies," *Computers & Electrical Engineering*, vol. 38, no. 5, pp. 1808-1819, 2012.
- [70] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Applied Soft Computing*, vol. 24, pp. 994-1012, 2014.
- [71] Y. Huang and T. Kechadi, "An effective hybrid learning system for telecommunication churn prediction," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5635-5647, 2013.
- [72] K. Coussement, "Improving customer retention management through cost-sensitive learning," *European Journal of Marketing*, vol. 3/4, pp. 477-495, 2014.
- [73] N. Lu, H. Lin, J. Lu and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659-1665, 2014.
- [74] A. Hanif and N. Azhar, "Resolving class imbalance and feature selection in customer churn dataset," in *Frontiers of Information Technology (FIT), 2017 International Conference on*, pp. 82-86, December 2017.
- [75] D. Do, P. Huynh, P. Vo and T. Vu, "Customer churn prediction in an internet service provider," in *Big Data (Big Data), 2017 IEEE International Conference on*, pp. 3928-3933, December 2017.
- [76] C. H. Lee, Y. H. Wang and A. J. Trappey, "Ontology-based reasoning for the intelligent handling of customer complaints," *Computers & Industrial Engineering*, pp. 144-155, 2015.
- [77] S. Moro, P. Rita and J. Coelho, "Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip," *Tourism management perspectives*, vol. 23, pp. 41-52, 2017.
- [78] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis and K. C. Chatzivasvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1-9, 2015.
- [79] M. A. H. Farquod, V. Ravi and S. B. Raju, "Churn prediction using comprehensible support vector machine: An analytical CRM application," *Applied Soft Computing*, vol. 19, pp. 31-40, 2014.
- [80] X. Zheng, S. Zhu and Z. Lin, "Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach," *Decision Support Systems*, vol. 56, pp. 211-222, 2013.
- [81] K. Kim, C. H. Jun and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6575-6584, 2014.
- [82] S. Pan, V. Giannikas, Y. Han, E. Grover-Silva and B. Qiao, "Using customer-related data to enhance e-grocery home delivery," *Industrial Management & Data Systems*, vol. 9, pp. 1917-1933, 2017.
- [83] Z. You, Y. W. Si, D. Zhang, X. Zeng, S. C. Leung and T. Li, "A decision-making framework for precision marketing," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3357-3367, 2015.
- [84] Z. P. Fan, Y. J. Che and Z. Y. Chen, "Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis," *Journal of Business Research*, vol. 74, pp. 90-100, 2017.
- [85] T. Verbraken, F. Goethals, W. Verbeke and B. Baesens, "Predicting online channel acceptance with social network data," *Decision Support Systems*, vol. 63, pp. 104-114, 2014.
- [86] Z. Y. Chen, Z. P. Fan and M. Sun, "A multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations," *European Journal of Operational Research*, vol. 255, no. 1, pp. 110-120, 2016.
- [87] M. Corazza, G. Fasano and F. Mason, "An Artificial Neural Network-based technique for on-line hotel booking," *Procedia Economics and Finance*, vol. 15, pp. 45-55, 2014.
- [88] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang and J. Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences*, vol. 235, pp. 117-129, 2014.
- [89] C. Yin, Y. Guo, J. Yang and X. Ren, "A new recommendation system on the basis of consumer initiative decision based on associative classification approach," *Industrial Management & Data Systems*, no. just-accepted, pp. 00-00, 2017.
- [90] J. Farhat and M. Owayjan, "ERP Neural Network Inventory Control," *Procedia Computer Science*, vol. 114, pp. 288-295, 2017.
- [91] L. Wang, R. Gopal, R. Shankar and J. Pancras, "On the brink: Predicting business failure with mobile location-based checkins," *Decision Support Systems*, vol. 76, pp. 3-13, 2015.
- [92] C. Wen, "8 Machine Learning CRMs That Are Moving The Industry Forward," GetCRM, 07 April 2017. [Online]. Available: <https://getcrm.com/blog/machine-learning-crm/>. [Accessed 15 April 2018].
- [93] R. Alt and O. Reinhold, "Social Customer Relationship Management (Social CRM): Application and Technology," *Business & Information Systems Engineering*, vol. 4, no. 5, pp. 287-291, 2012.
- [94] K. N. Lemon and P. C. Verhoef, "Understanding customer experience throughout the customer journey," *Journal of Marketing*, vol. 80, no. 6, pp. 69-96, 2016.
- [95] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," *arXiv preprint arXiv:1709.02779*, 2017.
- [96] K. Paulissen, K. Milis, M. Brengman, J. Fjermestad and J. Nicholas C. Romano, "Voids in the Current CRM Literature," in *Proceedings of the 40th Hawaii International Conference on System Science*, Hawaii, 2007.