

Departamento de Ciências e Tecnologias da Informação

***Machine learning* no processo de risco de crédito das instituições
bancárias**

André Luiz Monteiro

Dissertação submetida como requisito para obtenção do grau de

Mestre em Sistemas de Apoio à Decisão

Orientador(a):
Doutor, Rui Gonçalves,
ISCTE-IUL

Coorientadora:
Doutora Elsa Cardoso,
ISCTE-IUL

Setembro, 2019

Agradecimentos

Primeiramente, gostaria de agradecer a Deus, por sempre me manter no caminho certo durante este projeto com saúde e forças para chegar até o fim.

A filha Ana Luiza Feres Monteiro, que com seu jeito meigo e carinhoso foi sempre minha motivação e alegria.

A minha esposa Sofia José Feres, pelo seu esforço pessoal para estar ao meu lado em todos os momentos possíveis. Agradeço também pela sua compreensão deste desafio e as várias horas em que estive ausente.

A minha mãe Janilda Monteiro, pelo apoio que sempre me deu durante toda a minha vida.

A minha amiga de universidade e Banco do Brasil, Danielle Sandler, pelas valiosas contribuições dadas, força e compreensão.

A Prof. Dra. Elsa Cardoso, pela oportunidade e confiança que tornaram possível a realização deste sonho.

Ao meu orientador Dr. Rui Gonçalves pelo auxílio na execução deste trabalho, e agradeço a esta universidade, docentes e administração, por proporcionarem um ambiente de qualidade e excelência do ensino.

Ao Banco do Brasil, pelo apoio financeiro e institucional.

Eu agradeço com um forte abraço, a todos os amigos que de alguma forma fizeram parte dessa jornada.

Resumo

Uma vez que o sistema económico mundial se encontra em constante mudança, o estudo do risco de crédito tem uma grande importância para as instituições bancárias. Por estar associado a possíveis perdas que impactam o mercado financeiro, o processo de análise de crédito deve ser contínuo e progressivo.

O atraso nos pagamentos de negócios tornou-se uma tendência, especialmente após as recentes crises financeiras. Desse modo, os bancos devem minimizar dívidas, analisar individualmente os créditos, agir com rapidez e se proteger de não pagamentos.

Na mesma conjuntura, *machine learning* é uma tecnologia emergente para a construção de modelos analíticos, faz com que as máquinas aprendam com os dados. Com isso, efetuam análises preditivas de maneira mais rápida e eficiente. Fazer análises preditivas é muito importante e possui uma ampla gama de atuação para os bancos. Como, por exemplo:

- Identificação dos melhores fatores de risco a serem utilizados na antecipação de crédito a clientes;
- Obediência dos dispositivos legais;
- Qualidade de dados;
- Detecção de fraudes.

Na criação de uma pontuação de risco de crédito bancário, automatizada, robusta e eficaz, *machine learning* vai ajudar na previsão da capacidade de crédito do cliente com mais precisão.

O objetivo é analisar as diferentes abordagens de gestão de risco de crédito. Para tal, recorre-se a revisão de literatura de tópicos importantes, em destaque a *machine learning*, e ao uso de questionários.

Os principais resultados mostraram que o uso de *machine learning* no risco de crédito bancário, ainda está em fase inicial. A maioria dos bancos já reconhece os valores que esta tecnologia pode proporcionar. Com base nesses resultados, os bancos que são tão sensíveis a mudanças, têm que sair do âmbito da teoria e investir em pequenos projetos. Só assim esta tecnologia provará a sua capacidade de melhoria, e transmitir a confiança necessária para este sector.

Palavras-Chave: aprendizado de máquina; risco de crédito; instituições bancárias.

Abstract

As the global economic system is constantly changing, the study of credit risk is of great importance to banking institutions. Because it is associated with possible losses that impact the financial market, the process of credit analysis should be continuous and progressive.

Late business payments have become a trend, especially after the recent financial crises. Thus, banks should minimize debt, analyze individual credits, act quickly and protect themselves from non-payment.

At the same time, machine learning is an emerging technology for building analytical models, making machines learn from data. As a result, they carry out predictive analyses more quickly and efficiently. Predictive analysis is very important and has a wide range of activities for banks. For example:

- Identification of the best risk factors to be used in anticipating credit to customers;
- Compliance with legal provisions;
- Obedience of legal provisions;
- Data quality;
- Fraud detection.

In creating an automated, robust and effective bank credit risk score, machine learning will help predict the customer's creditworthiness more accurately.

The goal is to analyze the different approaches to credit risk management. To this end, a literature review of important topics is used, especially machine learning and the use of questionnaires.

The main results showed that the use of machine learning in bank credit risk is still at an early stage. Most banks already recognize the values that this technology can provide. Based on these results, banks that are so sensitive to change have to go beyond the scope of theory and invest in small projects. Only in this way will this technology prove its ability to improve and transmit the necessary confidence to this sector.

Keywords: machine learning; credit risk; banking institutions.

Índice

Agradecimentos	i
Resumo	ii
Abstract	iii
Índice.....	iv
Índice de Tabelas	vi
Índice de Figuras	vii
Lista de Abreviaturas e Siglas	ix
Capítulo 1 – Introdução	1
1.1. Enquadramento do Tema	1
1.2. Motivação e Relevância do Tema.....	2
1.3. Questões e Objetivos de Investigação	2
1.4. Abordagem Metodológica.....	3
1.5. Estrutura e Organização da Dissertação.....	3
Capítulo 2 – Revisão da Literatura.....	5
2.1. Risco de Crédito.....	5
2.1.1. Conceito de Risco	5
2.1.2. Conceito de Crédito	6
2.1.3. Conceito de Risco de Crédito	8
2.1.4. Gestão de Risco de Crédito	9
2.1.5. Spread	11
2.1.6. Medidas de Risco de Crédito.....	13
2.1.7. Principais Determinantes de Risco de Crédito	14
2.1.8. Processo de Gestão do Risco de Crédito	19
2.1.9. Modelos de Risco de Crédito.....	20
2.2. <i>Machine Learning</i>	24
2.2.1. Inteligência Artificial.....	24
2.2.2. <i>Big Data</i>	26
2.2.3. Conceito de <i>Machine Learning</i>	28
2.2.4. Categorias de <i>Machine Learning</i>	30
2.2.5. Algoritmos de <i>Machine Learning</i>	32
2.3. <i>Machine Learning</i> para Risco de Crédito.....	41
2.3.1. Vantagens	41
2.3.2. Estudos Comparativos	48
Capítulo 3 – Metodologia	53
Capítulo 4 – Análise e Discussão dos Resultados.....	55

4.1. Análise Bibliométrica	55
4.2. Questionário – Aos Técnicos de Crédito.....	60
4.2.1. Recolha de Dados	60
4.2.2. Resultados.....	60
4.3. Questionário – Aos Executivos e Diretores de Crédito	80
4.3.1. Recolha de dados	80
4.3.2. Resultados.....	80
4.4. Discussão.....	82
4.4.1. Questionário.....	83
4.4.2. Perfil do Entrevistado	83
4.4.3. Perfil das Instituições Bancárias.....	84
4.4.4. Conhecimento em <i>Machine Learning</i>	84
4.4.5. Conhecimento em <i>Machine Learning</i> para Risco de Crédito.....	86
Capítulo 5 – Conclusão.....	88
Bibliografia.....	91
Anexo	112
Anexo A – Questionário aos Técnicos de Crédito.....	112
Anexo B – Questionário aos Executivos e Diretores de Crédito.....	119
Anexo C – Outras Tabelas e Figuras	122

Índice de Tabelas

Tabela 1 – Linhas de pensamento sobre inteligência artificial.....	25
Tabela 2 – Características de big data e alguns dos principais autores	27
Tabela 3 – Categorias de machine learning.....	31
Tabela 4 – Vantagens de ML no processo de aprovação de crédito.....	42
Tabela 5 – Vantagens de ML no processo de acompanhamento de crédito.....	44
Tabela 6 – Vantagens de ML no processo de recuperação de crédito.....	45
Tabela 7 – Comparativo dos principais modelos de ML para crédito - 1996 a 2019.....	49
Tabela 8 – Grelha de pontuação para avaliação dos modelos de ML para crédito	51
Tabela 9 – Classificação da amostra de modelos de ML para crédito	52
Tabela 10 – Outros autores importantes	56
Tabela 11 – Principais Artigos	57
Tabela 12 – Análise de Likert para as afirmações de ML	78
Tabela 13 – Análise de Likert para as afirmações de ML para risco de crédito	79
Tabela 14 – Percentual de concordância nas afirmações de ML.....	85
Tabela 15 – Percentual de concordância nas afirmações de ML para RC	86
Tabela 16 – Ranking dos algoritmos de ML - Questionário	87
Tabela 17 – Estatística de preenchimento do questionário.....	122

Índice de Figuras

Figura 1 – Evolução da inteligência artificial e machine learning.	25
Figura 2 – Machine learning e deep learning no contexto da inteligência artificial.	29
Figura 3 – Detalhamento dos resultados dos modelos da amostra.	51
Figura 4 – Percentual de preferência de modelos de ML.	52
Figura 5 – Principais autores de machine learning para crédito.	55
Figura 6 – Principais termos de machine learning para crédito.	58
Figura 7 – Principais termos de machine learning para crédito por ano.	59
Figura 8 – Correlação dos principais termos de machine learning para crédito.	59
Figura 9 – Distribuição das respostas do questionário aos técnicos de crédito.	61
Figura 10 – Percentual de respostas por banco.	61
Figura 11 – Desempenho de respostas por banco.	61
Figura 12 – Tipo de cargo que exerce o entrevistado.	62
Figura 13 – Faixa de idade do entrevistado.	62
Figura 14 – Nível de graduação do entrevistado.	63
Figura 15 – Tempo de empresa do entrevistado.	63
Figura 16 – Você participa de alguma atividade relacionada a ML, Big Data ou TI? ...	64
Figura 17 – Quantidade de funcionários das empresas dos entrevistados.	64
Figura 18 – Volume de negócio das empresas dos entrevistados.	65
Figura 19 – Volume de negócio em crédito das empresas dos entrevistados.	65
Figura 20 – <i>Machine learning</i> trará benefícios aos negócios de sua empresa?	66
Figura 21 – Machine learning estará presente em todos os novos produtos e serviços? ...	66
Figura 22 – Considera que sua empresa atua em machine learning?	67
Figura 23 – Pensa em investir em pelo menos um projeto de ML até 2020?	67
Figura 24 – Falta de profissionais em ML é um dos problemas da sua empresa?	67
Figura 25 – A maioria dos executivos compreende os benefícios de ML?	68
Figura 26 – Os executivos acreditam que ML é um processo longo e lento?	68
Figura 27 – Precisam melhorar os processos a nível de dados para depois usar ML? ...	69
Figura 28 – Terceirização de TI para ML é uma das estratégias da sua empresa?	69
Figura 29 – Qual a área preferida para desenvolvimento em projetos de ML?	70
Figura 30 – O setor de risco de crédito já se utiliza de ML ou pensa em usar?	70
Figura 31 – Análise cruzada da atuação em Machine Learning por cargo.	71
Figura 32 – Análise aprofundada do uso de Machine Learning por cargo.	71
Figura 33 – ML pode auxiliar os analistas nas análises e concessões de crédito?	72
Figura 34 – Para gerentes e técnico, ML auxilia as análises e concessões de crédito? ..	72
Figura 35 – ML melhora os índices de incumprimento dos clientes?	73
Figura 36 – Para gerentes e técnicos, ML melhora os índices de incumprimento?	73
Figura 37 – ML auxilia na análise qualitativa dos C's do crédito.	73
Figura 38 – Machine Learning na análise qualitativa do crédito por tipo de cargo.	74
Figura 39 – Os RL, solvabilidade e rendibilidade são melhorados por ML?	74
Figura 40 – ML nos RL, solvabilidade e rendibilidade por tipo de cargo.	75
Figura 41 – Em que áreas do crédito já se utiliza ou pretende utilizar ML?	75
Figura 42 – Quais são os maiores desafios para se implementar ML em RC?	76
Figura 43 – ML trará mais eficácia para os negócios na ideia dos gerentes e técnicos. ...	77
Figura 44 – Likert - ML para RC por banco.	79
Figura 45 – Likert - ML para RC por nicho. Fonte: Elaboração própria.	79
Figura 46 – Validação do perfil das IB's pelos executivos ou diretores de crédito.	80
Figura 47 – Validação da utilização de Machine Learning em risco de crédito.	80
Figura 48 – Validação das questões de Machine Learning em risco de crédito.	81

Figura 49 – Desafios de ML em risco de crédito na visão dos executivos ou diretores.	82
Figura 50 – Percentual de respostas por nicho.	122
Figura 51 – Quais áreas do crédito para utilização de ML por nichos?	123
Figura 52 – Desafios de ML em RC por nichos.	123
Figura 53 – Os algoritmos de ML e eficácia para os negócios por nichos.	124
Figura 54 – A perspectiva de uso de ML em RC por escolaridade.	125
Figura 55 – ML podem auxiliar análises e concessões de crédito por escolaridade? ..	125
Figura 56 – ML e os índices de incumprimento por escolaridade.	125
Figura 57 – ML na análise qualitativa do crédito por escolaridade.	125
Figura 58 – ML nos RL, solvabilidade e rendibilidade por escolaridade.	125
Figura 59 – Quais áreas do crédito para utilização de ML por escolaridade?	126
Figura 60 – Desafios de ML em RC por escolaridade.	126
Figura 61 – Os algoritmos de ML e eficácia para os negócios por escolaridade.	127

Lista de Abreviaturas e Siglas

AI – *Artificial Intelligence*

AM – *Análise Multicritério*

APB – *Associação Portuguesa de Bancos*

BAS – *Bid-ask spread*

BCB – *Banco Central do Brasil*

BCBS – *Basel Committee on Banking Supervision*

BP – *Banco de Portugal*

CBS – *Corporate bond spread*

CDS – *Credit Default Swaps*

COPOM – *Comitê de Política Monetária*

COSO – *Committee of Sponsoring Organizations of the Treadway Commission*

CSFP – *Credit Suisse Financial Products*

DA – *Discriminant analysis*

DM – *Default model*

DP – *Dirichlet process*

DPMM – *Dirichlet process mixture model*

DT – *Decision trees*

EAD – *Exposure at default*

EDF – *Expected default frequency*

EDL – *Expected default loss*

EFQM – *European Foundation for Quality Management*

EL – *Expected Loss*

EUA – *Estados Unidos da América*

FERMA – *Federation of European Risk Management Associations*

FMDP – *Factored Markov decision processes*

GA – *Genetic algorithm*

GMM – *Gaussian mixture model*

HMM – *Hidden Markov model*

IB – *Instituição Bancária*

LGD – *Loss Given Default*

M – *Maturidade (Maturity)*

MAR – *Modelo de Avaliação de Riscos*

ML – *Machine learning*

MTM – *Mark-to-market*

NB – *Naïve Bayes Classification*

OASPRD – *Option-Adjusted Spread*

PD – Probabilidade de *Default*

R – *Recovery Rate*

RB – Rede Bayesiana

RC – Risco de crédito

RL – Rácios de liquidez

RLog – Regressão logística

RLin – Regressão logística

RNA – Rede neural artificial

SELIC – Sistema Especial de Liquidação e de Custódia

SVM – *Support Vectors Machine*

TI – Tecnologia da Informação

TSPRD – *Spread trade*

UE – União Europeia

UL – *Unexpected Loss*

YSPRD – *Yield Spread*

ZSPRD – *Z-Spread*

Capítulo 1 – Introdução

1.1. Enquadramento do Tema

Em época de instabilidade financeira mundial, principalmente após a concessão desregulada de crédito — crise do *subprime*, é fundamental que as instituições bancárias efetuem as suas análises com mais critérios e rigor. É importante que prevejam o comportamento dos seus clientes e adequem o crédito antes de autorizar um empréstimo.

Toda a concessão de crédito está relacionada a um risco. Portanto, o risco de crédito tem uma relação direta com a taxa de incumprimento.

No início do século XX, as escolhas dos proponentes de crédito eram baseadas no julgamento de um ou mais analistas (Thomas, Crook, J., & Edelman, D., 2017). Assim sendo, as decisões de aprovação de crédito eram subjetivas. Às vezes, numa mesma instituição, um pedido podia ser ou não aprovado, dependendo do analista que estava a julgá-lo.

Já Taleb (2007), na sua obra “O Cisne Negro” classifica os riscos desde o Cisne Branco (previsão alta e baixo impacto) até o Cisne Negro (previsão baixa e alto impacto). Nesse contexto, o pior risco é descrito como um acontecimento altamente improvável. Ser imprevisível, produzir grande impacto e explicar de forma menos aleatório e mais previsível do que aquilo que é na realidade. O autor também defende que não se pode evitar estes acontecimentos, mas sim, diminuir os seus efeitos negativos e parar de tentar prever eventos raros, admitindo que os nossos conhecimentos são limitados.

Desde então, muitos processos relacionados ao risco de crédito bancário foram evoluídos. Frente a massificação da *internet*, vemos o crescimento exponencial do volume e da variedade de dados partilhados na rede. Uma nova era está a surgir e nela torna-se imprescindível o uso de poderosas ferramentas computacionais — tais como nuvens e elementos de *big data*.

Atualmente, *machine learning* é o domínio de TI que mais contribui para problemas de previsão de negócios. É considerada a tecnologia ideal para (Ngai et al., 2011; Ian & Eibe Frank, 2011):

- Entender o padrão de transações bancárias;
- Identificar os clientes através de dados;
- Distinguir entre uma ação normal de uma fraude.

Neste sentido, toda a melhoria que *machine learning* proporcione sobre os processos de riscos de crédito, por envolver grandes quantias de dinheiro, gera uma economia enorme para as instituições bancárias.

1.2. Motivação e Relevância do Tema

Diante da crescente criação e partilha de dados, são gerados milhares de informações por segundo. No entanto, esses dados estão desestruturados, em diversas origens, com toda a espécie de atividade. Isto posto, dados e informações só terão valor se as instituições e órgãos públicos:

- Analisar e compreender os dados;
- Integrar os seus sistemas;
- Identificar tendências de eventos;
- Prever comportamentos.

Na sua natureza, dados digitais são desestruturados, muito ricos, porém, difíceis de explorar e entender (Tecnologia e Segurança, 2012). Com o rápido desenvolvimento digital, as empresas já têm dificuldades em utilizar somente os sistemas tradicionais de gestão de conteúdos e de tomada de decisão. Uma vez que os mesmos deixaram de ser capazes de oferecer respostas ágeis e precisas. A geração de dados em alto volume, velocidade e variedade exige a utilização de novas tecnologias que precisam coexistir com os tradicionais sistemas de registo das corporações.

O desafio das instituições bancárias não é diferente. Há uma imensidão de dados, diversas tecnologias, modelos, processos e comportamentos que precisam ser remodelados para enquadramento na nova era digital.

A grande contribuição deste estudo está relacionada as vantagens do uso das ferramentas de *machine learning* para gestão de risco de crédito. Uma vez que foi verificado que a área de crédito gera retornos financeiros significativos e *machine learning* ainda é uma tecnologia pouco explorada pelas instituições bancárias.

1.3. Questões e Objetivos de Investigação

O objetivo desta pesquisa é transmitir ao Banco do Brasil S.A., conhecimentos a respeito da utilização atual da indústria bancária, em especial a *machine learning* para risco de crédito.

De maneira mais específica é responder as seguintes questões:

- O que *machine learning* agrega de valor para o risco de crédito?
- Quais os fatores de sucesso a ter em conta para implementação em sistemas de *machine learning* no risco de crédito?
- Para as instituições bancárias quais os maiores desafios para a implementação de *machine learning* para risco de crédito?

1.4. Abordagem Metodológica

Inicialmente, utilizou-se a metodologia de pesquisa exploratória em livros, artigos acadêmicos e artigos de revistas atuais e especializadas, com intuito de adquirir familiaridade com os principais conceitos sobre o tema. Em paralelo, elaborou-se um questionário para levantamento de como as instituições bancárias estão trabalhando, tanto no âmbito de *machine learning* quanto de processos relacionados à risco de crédito.

Num segundo momento, já com os conceitos do tema mais apurados, fez-se uma pesquisa descritiva através de uma análise mais detalhada e minuciosa do objeto do estudo.

Cabe destacar que os dados utilizados são procedentes das respostas dadas ao primeiro questionário elaborado – cuja amostra limitou-se aos colaboradores de 12 bancos. Dentre eles estão bancos brasileiros, portugueses e internacionais (EUA, Espanha, França, Reino Unido e China).

Por fim, refinou-se, através dos resultados do primeiro questionário e a análise descritiva do tema, um segundo questionário com intuito de validar as informações obtidas, junto aos executivos ou diretores de crédito.

1.5. Estrutura e Organização da Dissertação

O presente estudo está organizado em cinco capítulos.

O primeiro capítulo introduz o tema da investigação e seus objetivos. Além disso, uma breve descrição da estrutura do trabalho.

O segundo capítulo reflete o enquadramento teórico, designado por revisão da literatura.

O terceiro capítulo é dedicado à metodologia utilizada no processo de recolha e tratamento de dados bem como aos métodos de análise utilizados.

O quarto capítulo apresenta a análise dos resultados obtidos.

No quinto e último capítulo apresentam-se as conclusões deste estudo bem como as recomendações, limitações e sugestões de trabalhos futuros.

Capítulo 2 – Revisão da Literatura

Neste capítulo abordou-se a noção básica dos dois principais temas deste estudo. Primeiramente, a noção de risco de crédito, e em seguida a de *machine learning*. Já na última secção do capítulo, combinou os dois assuntos.

2.1. Risco de Crédito

2.1.1. Conceito de Risco

A definição de risco não é um assunto recente. Markowitz na década de 50, já mencionava risco na sua obra “Moderna Teoria das Carteiras”. O autor defende que o risco de um ativo medido de forma isolada, torna-se diferente, quando esse é incluído numa carteira. Por consequência, o risco é a soma das variâncias individuais, e covariâncias entre os pares de ativos de uma carteira, considerando o peso de cada ativo da carteira. (Markowitz, 1952).

Na simbologia Hanzi da China, risco é representado por dois caracteres, sendo o primeiro associado a perigo e o segundo a oportunidade.

No âmbito da gestão, o crédito adapta-se bem ao conceito de risco. As instituições aumentam e materializam as suas oportunidades de negócio através do crédito. Por outro lado, existe sempre a ameaça do devedor de não pagar. Se esse risco se torna elevado e generalizado, a desconfiança recai a instituição, que pode levá-la a uma crise financeira (P. Carvalho, 2009).

No Michaelis Dicionário Brasileiro da Língua Portuguesa, risco é definido como a possibilidade de perigo. É a probabilidade de prejuízo ou de insucesso em determinado empreendimento, projeto, coisa, etc. em razão de acontecimento incerto, que não depende da vontade dos envolvidos. (Michaelis, 2015)

Já no dicionário Oxford University Press, o risco é a possibilidade de perda em termos absolutos ou em relação às expectativas. (Oxford University Press, 2005)

Em termos gerais, o risco é definido como a probabilidade de ocorrências de eventos divergentes face ao que era esperado (E. S. Silva et al., 2013).

O COSO considera risco como a possibilidade de que um fato aconteça e impacte negativamente o cumprimento dos objetivos (COSO, 2007). Em 2017, evoluindo sobre o

tema, ressalta a importância de se considerar o risco tanto no processo de definições das estratégias como na melhoria do desempenho (COSO, 2017).

Para Holton (2004), é importante diferenciar interpretações objetivas que são probabilidades reais, descobertas por lógica ou por análises estatísticas, e interpretações subjetivas que se baseiam em crenças humanas. As proposições têm probabilidades intrínsecas de serem verdadeiras ou falsas, podem depender da falta de informação. Este autor trata o risco como exposição a uma proposição que é incerta, e que a variância é muitas vezes um representante do risco. Ademais, relata que nunca se pode definir o risco. No máximo, definir a nossa percepção sobre ele (Holton, 2004).

O risco traduz a incerteza e o seu impacto. A incerteza, por si só, não conduz ao risco. É a incerteza aliada ao impacto das consequências que levam a situações de risco. Os riscos podem, ainda, ser considerados endógenos quando, até determinado ponto, estão sob controlo da organização. Caso contrário, denominam-se exógenos, como, por exemplo, os desastres naturais (Hull, 1992).

No contexto bancário os tipos de riscos podem ser distinguidos de acordo com a sua natureza (M. P. Amaral, 2015):

- Risco financeiro: quando o risco está diretamente relacionado aos ativos e passivos monetários da instituição;
- Risco não financeiro: quando o risco resulta de circunstâncias externas (fenómenos sociais, políticos ou económicos) ou internas (recursos humanos, tecnologias, procedimentos e outros) à instituição.

O BP (2007), através do documento denominado por MAR – Modelo de Avaliação de Riscos, identifica nove categorias de riscos. Dentre os de cunho financeiro destacam-se: o risco de crédito, risco de mercado, risco de taxa de juro e risco cambial; Já entre os não-financeiro: o risco operacional, risco dos sistemas de informação, risco de estratégia, risco de *compliance* e risco de reputação (Banco de Portugal, 2007).

Visto que o risco de crédito figura como o principal risco bancário e o mais difícil de quantificar (Morais, 2011), é nele que este estudo focará mais.

2.1.2. Conceito de Crédito

A palavra crédito deriva do termo latino *credere*, que significa acreditar. No Michaelis Dicionário Brasileiro da Língua Portuguesa, crédito é definido por confiança que

inspiram as boas qualidades de uma pessoa; confiabilidade, credibilidade. Boa reputação; apreço, deferência, estima. (Michaelis, 2015)

Crédito pode ter vários entendimentos. Ocorre quando uma pessoa física, pessoa jurídica ou instituições efetuam pagamentos a prazo. Quando há uma cessão temporária de fundos a terceiros, na qual se exige juros. Esses juros são definidos conforme o risco do devedor e a SELIC¹ praticada no mercado. (Ferreira, Celso, & Barbosa Neto, 2012; Ribeiro, Zani, & Zanini, 2009).

No contexto bancário, crédito foi definido pela APB (2015) sendo: “um direito que o banco adquire, através de uma entrega inicial em dinheiro (real ou potencial) a um cliente, de receber desse cliente o valor em dívida, em datas futuras, uma ou várias prestações em dinheiro cujo valor total é igual ao da entrega inicial acrescida do preço fixado para esse serviço” (juros e comissões) (APB, 2015).

O Banco Popular do Brasil, através do sítio web do Banco do Brasil, define crédito como: “um termo que traduz confiança, e deriva de “crer”, acreditar em algo ou alguém. O crédito, sob o aspeto financeiro, significa dispor a um tomador, recursos financeiros para fazer frente a despesas ou investimentos, financiar a compra de bens, etc.” (Banco Popular do Brasil, 2018).

Ainda, conforme A. Amaral, et al., (1997), o crédito é a troca financeira do credor, num dado instante, favorável ao devedor. O credor concorda com o risco de o montante ser delongado o seu pagamento para uma data futura com um juro de remuneração. É a confiança na realização do pagamento pelo devedor (A. Amaral et al., 1997).

Por fim, os autores definem o crédito ou a confiança como um atributo intrínseco ao indivíduo, a avaliação dessa percepção leva à caracterização de indivíduos ou empresas com diferentes níveis de confiança, favorecendo em muitos casos, o uso de medidas estatísticas de probabilidade, bem como a necessidade de análise periódica dos mesmos (Garcia & Giambiagi, 2010).

¹ SELIC é a taxa básica de juros definida pelo COPOM.

2.1.3. Conceito de Risco de Crédito

Por risco de crédito entende-se a perda de capital em caso de falha do tomador de crédito para reembolsar o montante total da dívida para recuperar o passivo (Gonen, Gonen, & Gargen, 2012).

Ademais, o risco de crédito pode ser atrelado a ocorrências de perdas associadas à não conformidade pelo tomador, ou contraparte de suas respectivas obrigações financeiras nos termos pactuados (BCB, 2017). Em outras palavras, o risco de crédito está diretamente relacionado ao risco de descumprimento.

Consoante com o Glossário do Banco de Portugal, risco de crédito é a possibilidade de o beneficiário de um dado empréstimo não ter capacidade financeira para pagar taxas de juros e / ou capital de dívida (Banco de Portugal, 2015).

Para o Banco do Brasil, risco de crédito é apresentado como a possibilidade de perda resultante da incerteza quanto ao recebimento de valores pactuados com tomadores de empréstimos, contrapartes de contratos ou emissões de títulos (Banco do Brasil, 2018).

Já para o Citi bank, risco de crédito se refere a queda da qualidade do crédito, ou a falha de um mutuário, contraparte, terceiro ou emitente, de honrar as suas obrigações financeiras ou contratuais (Citi bank, 2018).

Nesta mesma vertente, o Banco da China definiu o risco de crédito como o risco de que as partes, com quem o banco contratou, de não cumprir as suas obrigações (dentro e fora do balanço) (Bank of China, 2017).

As quatro metodologias mais usadas até os anos 90 para classificação de crédito foram (Altman & Saunders, 1998):

- Modelo linear de probabilidade;
- Modelo logit;
- Modelo probit;
- Modelo de análise discriminante.

Atualmente, além das técnicas estatísticas, existem as metodologias de *machine learning*. Enquanto as técnicas estatísticas tendem a ter os dados ajustados ao modelo — sendo geralmente necessário que os pesquisadores coloquem as estruturas do modelo — as técnicas de *machine learning* extraem automaticamente conhecimento a partir dos

dados, gerando modelos complexos que tendem a se ajustar aos dados (Z. Huang et al., 2004).

Segundo Dima e Orzea (2014), existem dois tipos de risco de crédito: risco *default* (incumprimento do crédito) e risco do *spread* de crédito. Por risco *default* entende-se com aquele que ocorre quando um emissor de dívida, devedor, não consegue honrar suas obrigações financeiras. Quando um devedor entra em *default*, o investidor geralmente incorre em uma perda igual ao montante devido pelo devedor, menos qualquer valor de recuperação que o investidor recupera como resultado da execução, liquidação ou reestruturação do devedor inadimplente. Já o risco de *spread* é o excesso de preço do governo ou a taxa livre de risco exigida pelo mercado para assumir uma certa exposição de crédito presumida. É importante notar que quanto maior o *rating* de crédito, menor o *spread* de crédito. Assim, o risco de *spread* de crédito é o risco de perdas financeiras resultantes de alterações no nível de *spreads* de crédito utilizados na tomada a mercado de um produto de renda fixa. As alterações nos *spreads* de crédito observados afetam o valor da carteira e podem levar a perdas para os *traders* ou a um desempenho insuficiente para os administradores de carteira (Dima & Orzea, 2014).

Conclui-se assim, que todas as organizações enfrentam múltiplos riscos e que não podem evitá-los, se querem crescer (Neves & Quelhas, 2013). Além de tudo, conforme (Gaspar, 2014) já havia destacado, o risco de crédito é o mais relevante para o setor bancário.

2.1.4. Gestão de Risco de Crédito

Antes de aprofundar no conceito de gestão de risco para crédito, iniciou-se este item com duas ideias sobre gestão de risco de forma geral.

Na primeira, COSO (2017) destaca que as organizações continuarão a sofrer com um futuro repleto de volatilidade, complexidade e ambiguidade.

A gestão de risco será um fator importante na condução dos negócios para prosperar em tempos de mudanças constantes. Todas as entidades precisam apresentar respostas eficazes às mudanças. Isto é, ter agilidade na tomada de decisão, responder de forma coesa, ser capaz de se adaptar e se reposicionar. No objetivo de manter o alto nível de confiança em todo o processo. Além disso, destaca-se a importância de saber lidar com a proliferação de dados, reconhecer que estamos na era digital e alavancar inteligência artificial e automação. Ter a preocupação que a gestão de risco precisa ser eficiente e que

proporcione o maior benefício possível a organização, principalmente com relação aos custos do processo (COSO, 2017).

A segunda citação traz que a gestão de risco, no atual ambiente de negócios competitivo, é definida pela FERMA como “gerir as ameaças e oportunidades para os nossos negócios dentro de tolerâncias aceitáveis a riscos” e desempenhar um papel vital no sucesso de todas as empresas (FERMA, 2018).

Doravante, relacionado especificamente a gestão de risco de crédito citou-se a EFQM, que manifesta gestão de risco de crédito como: “Um processo sistemático e transversal à organização bancária que permite identificar, gerir e monitorizar risco de crédito de forma que a informação gerada possa ser utilizada para proteger e criar valor” (EFQM, 2005). É da materialização dos objetivos de lucro e de valor contido na estratégia da organização que se obtém a gestão do risco de crédito (P. Carvalho, 2009).

Enfim, a eficácia da gestão do risco de crédito, é uma determinante primordial para o sucesso de qualquer organização bancária (BCBS, 2000).

De seguida, sucintamente, descreveu-se o processo de gestão do risco de crédito (E.S. Silva et al., 2013):

- Definição dos objetivos globais pretendidos para a gestão de crédito;
- Identificação de todas as situações capazes de expor a organização ao risco de crédito.
- Análise do risco, ou seja, a avaliação e quantificação objetiva ou subjetiva desse risco de crédito.
- Recolha, trataria e análise de todas as informações obtidas sobre o cliente, bem como o contexto que os rodeia.
- Depois desta análise é que se inicia a relação com o cliente,
- Monitorização do risco de crédito.
- Conclusão do processo, somente quando da liquidação total da dívida.

Além disso, prever e gerenciar o risco de crédito, cada banco deve ser capaz de quantificar os fatores relevantes de risco de crédito, analisar o risco envolvido e monitorar permanentemente os fatores de risco de crédito (Konovalova, Kristovska, & Kudinska, 2016).

Segundo o Santander Totta, as relações bancárias estão baseadas na articulação do binómio risco e rentabilidade. Defende-se ainda que a gestão de riscos dos bancos, essencialmente, passa por: “identificar, medir, analisar, controlar, negociar e decidir relativamente aos riscos incorridos pela operativa do banco” (Banco Santander Totta, 2015). As instituições devem privilegiar uma gestão de riscos rigorosa, eficiente e equilibrada (Banco Mundial, 2014).

Os bancos devem adaptar seus processos de crédito às condições de mudanças na economia do país e no padrão de vida da população. Para isso, os métodos para quantificar e analisar o risco de crédito são de grande importância para o bom funcionamento das instituições bancárias (Konovalova, Kristovska, & Kudinska, 2016).

Por tudo isto, verifica-se que a gestão do risco de crédito é um importante desafio para as instituições bancárias. Sendo que o seu insucesso pode levar a sua falência. De modo a evitar isso, a gestão de risco de crédito deve ter como meta atingir os seguintes objetivos (Konovalova, Kristovska, & Kudinska, 2016):

- Identificar padrões comuns de comportamento económico dos clientes bancários;
- Formular um conjunto de requisitos diferenciados para os mutuários;
- Criar grupos de acordo com a sua especificidade;
- Determinar o apetite ao risco da pessoa que toma decisões sobre o montante;
- Estipular, individualmente, prazo e juros de um empréstimo a ser concedido.

2.1.5. Spread

Nesta secção, o objetivo central é compreender o conceito de *spread* para crédito. De forma simplificada, entende-se que o *spread* é a margem de lucro das instituições bancárias no ato de um contrato de crédito. Os *spreads* de crédito têm como objetivo gerar uma remuneração extra ao investidor pelo facto de o mesmo estar a assumir vários riscos, especificamente: o risco de incumprimento da empresa, a incerteza do tamanho das perdas, a liquidez, e os impostos (J. -Z. Huang & Kong, 2003).

Segundo Balbi (2006), *spread* é a diferença entre os juros médios, que os bancos pagam ao recolherem o capital, e juros médios que recebem ao aplicá-los. (Balbi, 2006). Já para Collin-Dufresne, Goldstein e Martin, a existência dos *spreads* de crédito visa remunerar o investidor em dois aspetos: risco de incumprimento e, conseqüentemente, às suas perdas esperadas (*expected default loss*). Estes dois fatores, no risco de crédito, são

conhecidos como *corporate bond spread*. Ainda segundo autor, havendo incumprimento, isto é, em caso de *default*, o investidor não recebe o total dos seus investimentos, apenas uma parcela daquilo que lhe foi prometido. (Collin-Dufresne, Goldstein, & Martin, 2001).

Além do exposto, tem o entendimento de Segal que diz: um *spread* pode ter vários significados em finanças. Basicamente, no entanto, todos se referem à diferença entre dois preços, taxas ou rendimentos. Ele subdivide o *spread* em (Segal, 2019):

- *Bid-ask spread*: é a lacuna entre oferta e os preços de venda de um título ou ativo, como uma ação, uma obrigação ou uma mercadoria.
- *Spread trade*: é a diferença numa posição de negociação. É a diferença entre uma posição curta (ou seja, vender) de uma comprada. Normalmente, são feitas com opções ou contratos futuros. Neste caso, os spreads são precificados como uma unidade ou como pares, em futuras trocas, para garantir a compra e venda simultânea de um título. Isso elimina o risco de execução em que uma parte do par é executada, mas a outra falha.
- *Yield Spread*: também conhecido de spread de crédito. Evidencia a diferença entre as taxas de retorno apreciada por duas empresas de investimentos. Essas empresas, geralmente divergem sobre a qualidade de crédito. Isso é denominado por alguns especialistas de crédito de rendimento de X sobre Y.
- *Option-Adjusted Spread*: reduzir o preço de um título e compará-lo ao de mercado, a *yield spread* tem que ser juntada a uma curva de juros de referência. Esse preço reduzido é reconhecido de *spread* ajustado pela opção.
- *Z-Spread*: é o *spread* resultado das curvas de rendimento de juros de volatilidade zero. Tem como objetivo reduzir com o cronograma de fluxo de caixa originado para alvejar o seu preço de mercado. O *Z-Spread* é principalmente empregado para títulos lastrados em hipotecas. Todavia, pode ser utilizado em CDS (*Credit Default Swaps*) para medir o spread de crédito.

Enfim, as instituições bancárias visam a obtenção de lucro através de empréstimos aos seus clientes. Para tanto, tem de se financiar junto aos seus acionistas, clientes de depósitos à ordem ou a prazo, outros bancos, ou investidores. Em fazendo isso, terá que suportar este custo com essas fontes de financiamentos, em forma de taxa de juros.

2.1.6. Medidas de Risco de Crédito

Depois de tratar sobre spread passou-se para a relação de medidas de risco de crédito segundo (P. Carvalho, 2009), (Santos, 2011) e reforçado pelo (Financial Stability Institute, 2010):

A probabilidade de *default* - PD é a perspectiva de um cliente não honrar com as suas obrigações de pagamentos, num determinado tempo. Reflete o perfil de risco do cliente por transação e é obtido a partir de modelos de *rating* ou *scoring*.

Processos de exposição ou *exposure at default* - EAD é o valor em dívida pelo cliente, à instituição, no momento do incumprimento. No seu cálculo deve ser considerado a característica do crédito atribuído.

A taxa “R” ou *Recovery Rate* é a percentagem do montante de crédito concedido que a instituição financeira recupera, em caso de ocorrer *default*.

Por *Loss Given Default* - LGD entende-se como o valor que a instituição perde efetivamente. É quando um cliente entra em incumprimento e pode ser definida, através da taxa de recuperação ($R = 1 - LGD$). São as perdas previsíveis numa operação de concessão de crédito no caso de a contraparte entrar em incumprimento. Normalmente, apurada em percentagem da EAD.

A Maturidade - M equivale à medida efetiva do crédito, sendo uma proporção ponderada da vida da operação de crédito. É a percentagem do capital pago em cada ano, ponderada pelo ano a que diz respeito.

Neste ponto é importante reconhecer que as perdas na carteira de crédito são de difícil previsão no médio prazo. No entanto, pode-se estimar de perdas esperadas (*Expected Loss* - EL) que é a perda média de uma determinada carteira.

As ações das instituições bancárias não estão imunes as condicionantes externas que não se controla. Essa situação pode afetar o seu negócio e são denominadas como perdas não esperadas (*Unexpected Loss* - UL).

Para encerrar esta secção, é importante relatar que as instituições bancárias estão autorizadas a terem as suas fórmulas de classificação de risco de crédito. Por este motivo, estas perseguirão critérios mais rígidos de avaliação que proporcionarão mais segurança e maior transparência ao mercado.

Para fins acadêmicos os autores consideraram a metodologia IRB para o cálculo do risco de crédito, através da seguinte equação:

$$EL = PD * LGD * EAD.$$

Com isso, finalizou-se a parte do estudo que tratou das medidas de risco de crédito e a seguir deu-se início aos determinantes do risco de crédito.

2.1.7. Principais Determinantes de Risco de Crédito

Nesta seção segue a relação de alguns dos principais determinantes relacionados ao risco de crédito, bem como sua influência em algumas crises financeiras.

Porém, antes é necessário lembrar sobre algumas definições de correlação em estatística, que serão mencionadas em todas as determinantes:

- a. Correlação positiva: é quando existe uma relação entre as variáveis, e elas tendem a mudar na mesma direção, ou seja, se uma variável aumenta, a outra sempre vai aumentar;
- b. Correlação negativa: é quando existe uma relação entre as variáveis, mas agora elas mudam em sentido oposto. Isto é, quando uma variável aumenta, a outra sempre vai diminuir;
- c. Correlação significativa: é quando existe uma probabilidade aceitável de “existir” uma correlação entre as variáveis.
- d. Correlação não significativa ou insignificante: é quando existe uma probabilidade aceitável de “não existir” uma correlação entre as variáveis.

Agora em relação as determinantes de crédito, iniciou-se com a oferta de moeda que é aquela que é definida pelo total de dinheiro em circulação ou que existe num país. Poudel relata nos seus estudos que o aumento da oferta de moeda é diretamente relacionado ao aumento da produtividade e lucro. Logo, isso reflete num estímulo para o investimento e consumo, assim que a receita aumenta. Além disso, possui uma relação inversa com as taxas de juros. Ou seja, quanto maior a oferta de moeda, menor é a taxa de juros, e por consequência, maiores são as oportunidades de se obter fundos mais baratos (Poudel, 2013).

Para complementar o estudo da determinante oferta de moeda, elencou-se sobre algumas crises e as perspectivas de alguns autores com relação a sua influência:

- De acordo com os autores, a oferta de dinheiro estava correlacionada negativamente ao risco de crédito na crise financeira da Roménia (Bucur & Dragomirescu, 2014);
- Já para Mileris, na crise dos 22 Países da União Europeia, a determinante de oferta de moeda teve uma participação significativa (Mileris , 2012);
- Enquanto no estudo do Poudel sobre a crise do Nepal, esta determinante não foi significativa (Poudel, 2013).

A segunda determinante deste estudo foi relacionada ao crescimento de crédito. De acordo com Castro, se o crédito cresce rapidamente, o risco de incumprimentos poderá aumentar no futuro. Com relação à perspectiva de alguns autores segue a influência do crescimento do crédito em algumas crises:

- Nas crises de alguns países da Europa (Grécia, Irlanda, Portugal, Espanha e Itália) a determinante de crescimento foi correlacionada positivamente com a crise (Castro, 2013).
- Já para o Washington a correlação desta determinante na crise no Quênia foi negativa (Washington, 2014).

O índice de confiança do consumidor é o determinante de risco de crédito que expressa o grau de otimismo do consumidor em referência à sua condição pessoal e a do país no curto e médio prazo (Mendicino & Punzi, 2013). Na perspectiva dos autores sobre crise dos 22 países da União Europeia, a correlação desta determinante de confiança do consumidor teve uma influência significativa para a crise (Mendicino & Punzi, 2013; Mileris , 2012).

Outra determinante importante é o produto interno bruto que é a soma monetária de todos os bens e serviços finais praticados num país, durante um período. Durante as crises financeiras os empréstimos são contratados a taxas mais elevadas, logo, é de se esperar uma relação negativa entre o *spread* de crédito e crescimento económico. Com relação às crises os principais autores perceberam que:

- Na crise financeira da Austrália e USA o PIB teve uma correlação negativa (Ali & Daly , 2010).
- Na Eslovénia a determinante PIB foi considerada insignificante (Aver, 2008).

- Já para Castro nas crises da Grécia, Irlanda, Portugal, Espanha e Itália, sua correlação foi negativa com os Respetivos PIB (Castro, 2013);
- No entanto Para Mileris, o PIB foi sim uma determinante importante para as crises dos 22 Países da União Europeia (Mileris , 2012);
- Por fim, o PIB teve uma correlação negativa na crise o Quénia (Washington, 2014).

O índice de preços de ações é mais um determinante de risco de crédito. Este significa o índice que reflete de forma geral as condições financeiras das principais empresas do mercado de um país. A análise dessa determinante identifica informações úteis para prever o estado atual e futuro da economia (Castro, 2013). Como nas determinantes anteriores seguem algumas percepções de autores da influência desta determinante com relação às crises:

- Na crise da Eslovénia o autor indica uma correlação positiva desta determinante (Aver, 2008).
- Enquanto que para a crise da Grécia, Irlanda, Portugal, Espanha e Itália, o estudo do autor indicou uma correlação negativa (Castro, 2013);
- Já o estudo de Mileris afirmou que esta determinante foi significativa para a crise dos 22 Países da União Europeia (Mileris , 2012).

A determinante taxa de câmbio é um indicador de importância para identificação de instabilidade económica. Uma vez que as empresas estão cada vez mais globalizadas, então operam sob a influência constante da taxa de câmbio de moeda estrangeira. Se o câmbio aumenta, torna-se mais caro importar, o lucro diminui e o risco de incumprimento cresce. E por isso as instituições bancárias ficam mais cautelosas no apoio a empresas. A seguir segue a influência desta determinante em algumas crises:

- O autor reporta que a determinante taxa de câmbio não foi significativa na crise da Eslovénia (Aver, 2008);
- Já na crise da Roménia a taxa de câmbio foi indicada como correlacionada ao risco de crédito de forma negativa (Bucur & Dragomirescu, 2014).
- Nas crises dos países da Grécia, Irlanda, Portugal, Espanha e Itália, o autor identificou uma relação entre a determinante taxa de câmbio e risco de crédito, e esta foi considerada positiva. (Castro, 2013)
- Já para Mileris, na crise dos 22 Países da União Europeia, a taxa de câmbio foi significativa para o risco de crédito (Mileris , 2012).

- Na crise financeira do Nepal, o autor considerou que esta determinante foi correlacionada negativamente com o risco de crédito (Poudel, 2013).
- Por fim, Washington também encontrou uma correlação negativa entre taxa de câmbio e risco de crédito na crise no Quênia (Washington, 2014).

Outra determinante é a taxa de desemprego. Esta é explicada pelo número de desempregados sobre a quantidade da população ativa. Com relação às perspectivas de alguns autores, segue a influência do desemprego em algumas crises:

- Na crise da Eslovênia, o autor considerou que esta determinante foi correlacionada positivamente com o risco de crédito (Aver, 2008).
- Na crise da Romênia, Bucur e Dragomirescu também correlacionaram a taxa de desemprego como positiva (Bucur & Dragomirescu, 2014).
- Nos países — Grécia, Irlanda, Portugal, Espanha e Itália, a correlação também foi positiva com o risco de crédito pelo autor (Castro, 2013).
- No entanto, para Mileris na crise dos 22 Países da União Europeia a correlação foi apenas significativa (Mileris, 2012).

Mais uma importante determinante é o saldo da balança comercial de país. É definida pela diferença entre o montante das exportações e o montante das importações durante um período. Quanto a influência das taxas de exportações e importações os autores mencionaram:

- Aver na crise da Eslovênia declarou como não significativa a correlação dessas taxas no risco de crédito (Aver, 2008).
- Enquanto que Mileris verificou que apenas as importações tiveram uma correlação significativa na crise dos 22 Países da União Europeia (Mileris, 2012).

A taxa de inflação é a determinante de risco de crédito que significa um aumento do nível geral de preços. A inflação é importante para os bancos visto ser um indicador financeiro que afeta a sua eficiência. Altas taxas de inflação são normalmente associadas a elevadas taxas de juro, aumentam os custos dos empréstimos e, por consequência, agravam o risco de crédito.

- Na Eslovênia, o autor indicou que a correlação desta determinante não foi significativa ao risco de crédito na sua crise (Aver, 2008).

- Já na crise da Roménia, de acordo com o autor, a correlação foi de forma negativa entre a determinante e o risco de crédito (Bucur & Dragomirescu, 2014).
- Castro mencionou uma correlação bilateral para as crises da Grécia, Irlanda, Portugal, Espanha e Itália. Ou seja, a correlação entre o risco de crédito e a inflação pode ser tanto positiva quanto negativa (Castro, 2013).
- Nos 22 Países da União Europeia, o autor correlacionou a taxa da inflação com risco de crédito apenas como significativa (Mileris , 2012).
- Na crise do Nepal foi indicado pelo autor que a inflação e o risco de crédito estão correlacionados negativamente (Poudel, 2013).
- Igualmente ao ocorrido no Nepal, no Quénia o autor também correlacionou como negativa a influência da inflação ao risco de crédito na sua crise (Washington, 2014).

E por último, taxa de juro é o determinante que define o montante a ser cobrado. É expresso como uma percentagem do capital, por um credor a um devedor para o uso de ativos. Este indicador é importante porque o seu aumento pode desencadear o crescimento do custo do investimento, que gera um risco de incumprimento maior, e consequentemente, prejudica o risco de crédito.

- Na crise financeira da Austrália e USA, o autor identificou que a correlação dos juros não foi significativa ao risco de crédito (Ali & Daly , 2010).
- Também na Nigéria, no momento da sua crise, a correlação da taxa de juro foi insignificante ao risco de crédito de acordo com o autor (Arewa, Nwachukwu, & Owoputi, 2013).
- Já na crise da Eslovénia o autor correlacionou de forma positiva a taxa de juro ao risco de crédito (Aver, 2008).
- Enquanto que Bucur e Dragomirescu, na crise da Roménia, correlacionaram a taxa de juro ao risco de crédito, porém de forma negativa (Bucur & Dragomirescu, 2014).
- Nos países - Grécia, Irlanda, Portugal, Espanha e Itália, Castro constatou que a taxa de juro tem uma correlação positiva ao risco de crédito (Castro, 2013).
- Já na crise dos 22 Países da União Europeia a correlação foi declarada pelo autor com apenas significativa (Mileris , 2012).

- De acordo com o estudo de Poudel (2013) com relação a crise financeira do Nepal a correlação não foi significativa entre os juros e o risco de crédito (Poudel, 2013).
- E na crise financeira do Quênia, o autor verificou que a taxa de juro teve uma correlação positiva com o risco de crédito (Washington, 2014).

Encerraram-se aqui as explicações dos autores quanto aos fatores determinantes do risco de crédito e suas influências em algumas crises financeiras. Passou-se assim, a analisar o risco de crédito com o viés de seus processos.

2.1.8. Processo de Gestão do Risco de Crédito

Depois de efetuado os estudos das medidas e determinantes do risco de crédito é importante reportar como abordar ou gerir esse mesmo risco.

De acordo com Carvalho, um sistema robusto de gestão é aquele que não só garante a continuidade do negócio como evita a falência. Para isso ele descreve o processo de gestão de risco de crédito em três etapas fundamentais: aprovação, acompanhamento e recuperação (P. Carvalho, 2009).

Aprovação é a etapa responsável por receber as propostas de crédito (novas ou renovações) e posterior análise. Esta consiste tanto em analisar quantitativamente quanto qualitativamente o cliente. O resultado desta etapa é um parecer de recomendações de tomada de decisão, onde é indicado aprovado para concessão de crédito ou não.

Entende-se por acompanhamento a de monitorização contínua do crédito e do cliente. Tem como a sua principal função a de prevenir um possível incumprimento. Uma vez identificado uma situação problemática ou cliente com alta probabilidade de incumprimento, é nesta etapa que se elabora um planeamento de ações, sejam elas, com a equipa responsável pela recuperação, ou com o cliente. Além disso, é também nesta etapa que se faz a regularização de incumprimentos dos clientes que já estão identificados com crédito negativo.

Enfim, tem-se a etapa de recuperação que é responsável em acolher os clientes em situação de incumprimentos já registados. É quando se tenta, amigavelmente uma negociação com o cliente. Caso não se resolva a situação de incumprimento nesta etapa, o problema sai da esfera de crédito e passa a ser um processo judicial. Este, portanto, deverá ser conduzido pelos especialistas jurídicos da empresa.

2.1.9. Modelos de Risco de Crédito

Existem várias formas de categorizar os modelos de risco de crédito. Uma delas é dividi-los em modelos qualitativos ou quantitativos. Os modelos são qualitativos quando a análise é de julgamento subjetivo, baseada nas características do tomador (reputação, alavancagem financeira, volatilidade dos lucros, garantias, e do mercado (ciclo de negócios, nível de taxas de juros) (Saunders e Cornett, 2011 citado por Zamore et al., 2018). Enquanto, os quantitativos são aqueles que ocorrem através de uma pontuação de crédito, usada para determinar o PD ou para classificar os mutuários em vários grupos de risco padrão (Saunders e Cornett, 2011 citado por Zamore et al., 2018).

Outra forma de classificar os modelos de risco de crédito é categorizá-los de pontuação, de classificação ou híbridos. Modelos de pontuação ou *credit scoring* são aqueles cuja análise estatística da qualidade de crédito estima a probabilidade de incumprimento na forma de *score* (Caeiro citado por Lisboa & Miguel, 2018). Modelos de classificação ou *credit rating* são semelhantes ao anterior, porém, a análise da qualidade de crédito é através de uma classificação estabelecida com base no histórico financeiro do cliente em cumprir com as suas obrigações. (Caeiro citado por Lisboa & Miguel, 2018. Por fim, os modelos híbridos que são a combinação entre os modelos, de forma que compreenda uma vertente de análise qualitativa e outra quantitativa. São os que ajudam as empresas e bancos a medir a perda potencial que uma *exposer* de crédito não pode oferecer (Klieštik & Cúg, 2015).

De forma a detalhar, elencou-se a seguir alguns dos principais modelos existentes:

A. Modelo *Credit Metrics*

O modelo *Credit Metrics* foi publicado em 1997 pelo banco JP Morgan e está estritamente ligado ao “Risco de Degradação de Crédito”. É um modelo estatístico, estrutural, baseado em técnicas e métodos quantitativo, no qual sua definição é MTM (mark-to-market), podendo ter uma abordagem de simulação ou analítica, impulsionado pelos valores de ativos, de volatilidade constante, com retornos sobre ativos normais de variáveis múltiplas, cujos índices de recuperação são aleatórios (Saunders, 1999; Adamko, Klieštik, & Mišanková, 2014).

Este modelo tem a prerrogativa de não só efetuar a análise padrão, como também, verificar o risco de mudanças nos preços de obrigações (Kollar, Weissovalvana, &

Siekelova, 2015). O ponto principal deste modelo está em medir o valor esperado e mensurar mudanças não previstas dentro do total da carteira de crédito (Saunders, 1999).

As mudanças de valores devem ser aferidas através da possibilidade de transição de uma categoria de crédito para outra (por exemplo, migração de uma empresa ou título de dívida do rating AAA para o rating B), tal como, a probabilidade de incumprimento (Saunders, 1999).

Sua estrutura começa do risco independente para o risco conjunto (Gupton, Finger, & Bhatia, 1997). São dois os desafios principais:

- a inexistência de normalidade nos valores agregados da carteira;
- a variação da carteira de crédito é mais difícil do que uma que considere apenas o risco de mercado.

Este modelo tem como objetivo principal de determinar que o preço das ações da empresa é aplicado como valor dos seus ativos. Além disso, o incumprimento só será transcorrido na data de vencimento do título de dívida, quando os valores dos ativos da empresa forem menores que o pagamento esperado pelos devedores dos títulos (Gupton, Finger, & Bhatia, 1997).

B. Modelo *Credit Risk +*

Em 1996, o Modelo “*Credit Risk +*” desenvolvido pelo *Credit Suisse Financial Products* (CSFP) foi apresentado como modelo de risco de crédito. É um modelo estatístico do padrão do risco de crédito, com intensidade, baseado em técnicas e métodos quantitativos, no qual a sua definição é DM (*default model*). É de abordagem analítica e atuarial, impulsionado pelo índice de incumprimento esperado. Possui volatilidade variável, com suposição de independência ou correlação com índice de incumprimento esperado. Cujos índices de recuperação são constantes dentro da faixa de qualidade do crédito (Saunders, 1999; Han, 2018).

O “*Credit Risk +*” se diferencia de outros modelos por não precisar supor às taxas de incumprimento em sua análise de risco de crédito (Mun, 2004 citado por Lisboa & Miguel, 2018). Assume que cada devedor tem uma PD única e que não existe uma correlação nos padrões entre dois emissores, a menos que eles compartilham fatores de risco semelhantes (Gordy, 2000 citado por Zamore, et al., 2018). As taxas de recuperação não fazem parte da modelação do “*Credit Risk +*”, são consideradas como variáveis

externas ao modelo. As exposições utilizadas no modelo são valores que já foram descontados da taxa de recuperação esperada (Andrade, 2004).

O principal objetivo do modelo é a medição de perdas esperadas e inesperadas da carteira de crédito. Como no “*Credit Metrics*”, o modelo “*Credit Risk +*” prevê a qualidade dos ativos de crédito e é muito volátil às probabilidades de incumprimento (Gordy, 2000 citado por Zamore, et al., 2018).

Os devedores são classificados em faixas de qualidade de crédito, onde cada faixa está associada a uma probabilidade de ocorrência de *default* e uma volatilidade da taxa de incumprimento² (Han, 2018). O pagamento ou o *default* ocorre apenas na data de vencimento do contrato de crédito.

Este modelo pressupõe quatro hipóteses (Han, 2018):

- Todo crédito individual apresenta apenas dois estados possíveis: com ou sem falhas;
- A probabilidade padrão de um crédito individual é baixa;
- A probabilidade de incumprimento para um grande grupo de tomadores é muito baixa;
- O número de incumprimento durante um período é independente de qualquer outro período.

Como principais vantagens apresenta-se (Han, 2018):

- Não exige grande quantidade de dados para a sua execução;
- Simples e fácil de se utilizar;
- Expressões de fórmulas matemáticas fechadas atrativas à implementação computacional
- Execução rápida

Já como limitações pode-se citar (Han, 2018):

- Não considera a existência do risco de mercado;
- Centraliza seus esforços nas perdas totais em detrimento das mudanças de valor;

² A volatilidade da taxa de incumprimento é o desvio padrão que seria observado numa carteira diversificada de crédito homogêneo.

- Não supõe migrações de *rating*. O risco de crédito de uma empresa durante a execução de um título é fixo;
- As técnicas de cálculo usadas não são necessariamente acessíveis a todos os utilizadores.

Por fim, a aplicação do “*Credit Risk +*” somente será válida se a probabilidade de incumprimento individual for baixa e os números diferentes de títulos de crédito forem altos (Aragão, Carvalho, & Barros, 2002).

C. Modelo *Credit Portfolio View*

O Modelo *Credit Portfolio View* foi desenvolvido por Wilson (1997) e proposto pela *McKinsey Consulting Company*. Está estritamente ligado ao “Risco de Incumprimento”. É um modelo do risco de crédito, no qual sua definição pode ser DM (*default model*) ou MTM. Possui uma abordagem de simulação, impulsionado por fatores macroeconómicos. É de volatilidade variável, com correlação com carregamentos dos fatores, cujos índices de recuperação são aleatórios (Saunders, 1999).

Este modelo tem como hipótese: a probabilidade de incumprimento, assim como a possibilidade de migração de qualidade de crédito. Por serem ligados à fatores macroeconómicos, pode-se afirmar que quando a economia piora, tanto a deterioração da qualidade de crédito como os índices de incumprimento aumentam. Por outro lado, quando a economia melhora, a qualidade de crédito cresce e o número de incumprimento diminui (Aragão, Carvalho, & Barros, 2002).

É uma combinação dos modelos “*Credit Risk +*” e “*Credit Metrics*”. Na primeira etapa, as probabilidades de *default* são estimuladas. Na segunda etapa, os *defaults* independentes — transições de *rating*, são amostradas, condicionadas às probabilidades simuladas, como em “*Credit Risk +*”. As probabilidades estimuladas estão vinculadas a variáveis macroeconómicas, bem como a valores semelhantes aos dos ativos que determinam os *defaults* individuais no modelo “*Credit Metrics*” (Witzany, 2017).

E por fim cita-se algumas limitações (Lueska, 2009):

- Necessidade de grande quantidade de dados macroeconómicos e de mercado (taxas de juros, preços de ações, spreads de crédito, preços de *commodities*).
- Eventos atípicos prejudicam as relações macroeconómicas existentes;

D. Modelo KMV

O Modelo KMV foi concebido pela *KMV Corporation*. O seu propósito é calcular a frequência esperada de incumprimento (EDF – *Expected Default Frequency*). Está associado a análise de ativos de crédito segundo a teoria de opções. A sua definição pode ser enquadrada em DM ou MTM. Possui uma abordagem analítica, impulsionado pelos valores de ativos. É de volatilidade variável, com retornos sobre ativos normais variáveis múltiplas, cujos índices de recuperação podem ser constantes ou aleatórios (Saunders, 1999).

O modelo de portfólio KMV combina as principais características da abordagem “*Credit Metrics*” e a metodologia KMV (*Expected Default Frequency*) (Witzany, 2017). Deriva das possibilidades de incumprimento e de transições de qualidade do crédito de uma empresa. A ideia geral é conceder um valor justo, uma vez que a empresa tomadora do crédito pode ser apontada como uma opção de venda (Saunders, 1999).

A principal vantagem que o KMV tem sobre modelos concorrentes é que a base do cálculo das possibilidades de incumprimento são os valores das ações da empresa (Caouette, Altman, Narayanan, & Nimmo, 2008). Por fim, a desvantagem é que ela é apropriada apenas para uma carteira de empréstimos ou títulos de empresas listadas publicamente, não para carteiras de venda a retalho ou de pequenas e médias empresas, em que os dados do mercado de ações geralmente não estão disponíveis. (Witzany, 2017).

Com isso, encerrou-se a parte relacionada ao primeiro tema importante de estudo – risco de crédito e passou-se por referenciar a literatura sobre *machine learning*.

2.2. *Machine Learning*

Não se pode introduzir o assunto *machine learning*, sem tratar de alguns termos que são usados para descrever este campo da ciência.

2.2.1. Inteligência Artificial

Inteligência artificial não é um tema novo. Relatam-se que Aristóteles, na época que almejava substituir a mão de obra escrava por objetos autónomos, foi um dos primeiros idealizadores de inteligência artificial. A AI se aprofundou, nos anos 50, com os pensadores: Herbert Simon e John McCarthy.

Na Figura 1 é apresentada a linha de evolução de inteligência artificial e *machine learning* ao longo dos tempos:

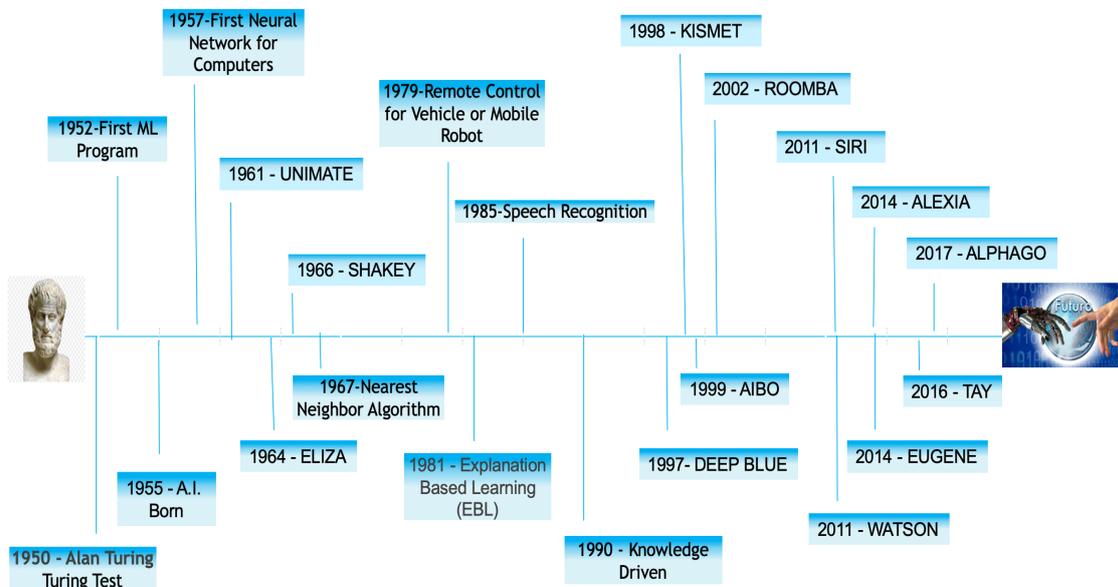


Figura 1 – Evolução da inteligência artificial e machine learning.

Fonte: adaptado de (Mardsen, 2017), (Thota, 2019).

Após a Segunda Guerra Mundial, definiu-se AI como o ramo da computação que estuda o comportamento inteligente (Luger G. , 2009). De outra forma, é a temática dos computadores praticarem ações que, atualmente, os homens realizam melhor (Rich, Knight, & Nair, 2009).

A AI evoluiu durante todos esses anos, todavia os conceitos ficaram entre os quatro princípios demonstrados na Tabela 1:

Tabela 1 – Linhas de pensamento sobre inteligência artificial

Pensando como um humano	Pensando racionalmente
<p>“O novo e interesse esforço para fazer os computadores pensarem(...) <i>máquinas com mentes</i>, no sentido total e literal (Haugeland, 1985).”</p> <p>“(Automatização de) atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado (Bellman,1978).”</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais (Charniak e McDermott, 1985).”</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir (Winston, 1992).”</p>
Agindo como seres humanos	Agindo racionalmente
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas (Kurzweil,1990).”</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas (Rich e Knight, 1991).”</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes (Poole et al.,1998).”</p> <p>“AI...está relacionada a um desempenho inteligente de artefactos (Nilson, 1998).”</p>

Fonte: (Russell & Norvig, 2013).

Os princípios I e III, concernem-se ao processo de pensamento e raciocínio. Enquanto, II e IV dizem respeito a comportamento. Os princípios I e II aferem o sucesso como fidelidade ao desempenho humano. Ao passo que, III e IV determinam o sucesso, comparando-o a um conceito ideal de inteligência — racionalidade. Um sistema é racional se “faz tudo certo”, com os dados que tem (Russell & Norvig, 2013). Depois de inteligência artificial, o outro termo importante de se saber é sobre *big data*.

2.2.2. *Big Data*

No mundo de hoje, onde todos estão sempre conectados, são gerados milhares de informações por segundo. É produzida uma imensidão crescente de dados desestruturados, de diversas origens e categorias. Por exemplo, o *Twitter* processa mais de 70 milhões de *tweets* por dia, gerando mais de 50 TB por dia (Cooper, 2019). Conforme estimativa, até 2022, haverá mais de 40 bilhões de dispositivos conectados (ABI Research, 2018). Este fenômeno, os teóricos padronizaram de *big data*.

Segundo o autor, *big data* são dados de imenso volume, alta velocidade e grande variedade. Posto isso, novos paradigmas de processamento deverão existir, para que: faça-se uso de grandes quantidades de dados, descobertas de intuições, melhorias na tomada de decisões e a otimização de processos (Mayer-Schönberger & Cukier, 2013).

O termo *big data* é definido como alto volume, alta velocidade e alta variedade de dados que exigem novos paradigmas de processamento para permitir a descoberta de insights, a melhoria na tomada de decisões e a otimização de processos. De acordo com esta definição, o *big data* não é caracterizado por métricas de tamanhos específicos, mas pelo fato de as abordagens tradicionais estarem lutando para processá-las devido a seu tamanho, velocidade ou variedade (Gartner Group, 2012).

Tal como, Hoe (2013) sinaliza que além das citadas pelo Gartner, *big data* também considera variabilidade e complexidade. Por variabilidade deduz a disponibilidade de informações sujeitas a ‘picos periódicos’ e a eventos. Já complexidade, há o desafio de gerir os dados. Isto é, integrar as suas fontes, parar de duplicar informações e atualizar dados (Hoe, 2013).

A capacidade de extrair valor do *big data* depende dos processos atrelados à análise de dados. A análise de dados envolve várias abordagens, tecnologias e ferramentas, como as de análise de texto, inteligência de negócios, visualização de dados e análise estatística (Jagadish, et al., 2014).

Na Tabela 2, tem-se um resumo destas características de *big data*, bem como a menção de alguns dos seus principais autores:

Tabela 2 – Características de big data e alguns dos principais autores

Características	Descrição	Referência
Volume	Grande quantidade e complexidade de dados.	(Agrawal, 2014)
	Produção de 2,5 exabytes de dados por dia, duplicados a cada 40 meses.	(McAfee & Brynjolfsson, 2012)
	Refere-se à grande quantidade de dados gerados a cada segundo. Pode ser em terabytes, zettabytes ou até mais.	(Budhathoki, Dasgupta, & Jain, 2018)
Velocidade	Exigência de resoluções em curto prazo ou em tempo real.	(Goldman, Kon, Pereira Junior, Polato, & Pereira, 2012)
	Coleta e processamento de dados em quase tempo real.	(McAfee & Brynjolfsson, 2012)
	Captura rápida e análise contínua de dados.	(Zikopoulos, Lightstone, Huras, & Sachedina, 2013)
	Refere-se à velocidade com que os dados estão a ser gerados ou produzidos. Pode-se analisar esses dados à medida que eles são gerados de forma muito eficiente.	(Budhathoki, Dasgupta, & Jain, 2018)
Variedade	Grande variedade de fontes e formas de dados com o desafio de encontrar os padrões úteis para o negócio.	(McAfee & Brynjolfsson, 2012)
	Diversidade de origens, formas e formatos de dados.	(Demchenko, Grosso, De Laat, & Membrey, 2013)
	Os dados vêm em vários arranjos, por exemplo, dados brutos, organizados, semiestruturados e não-estruturados.	(Memon, Soomro, Jumani, & Kartio, 2017)
	Indica a complexidade, que vem da grande heterogeneidade de dados.	(Cheng, Fang, Yang, & Cui, 2018)
Veracidade	Autenticidade, reputação de origem e confiabilidade dos dados.	(Demchenko, Grosso, De Laat, & Membrey, 2013)
	Pode ser definido como anormalidade nos dados. 40 a 50% do tempo é gasto na limpeza dos dados.	(Budhathoki, Dasgupta, & Jain, 2018)
	Os dados podem estar com ruídos, imprecisos e redundantes.	(Cheng, Fang, Yang, & Cui, 2018)
Valor	Depende da extração de valor económico do volume massivo.	(Cheng, Fang, Yang, & Cui, 2018)
	Dados com significado para os negócios, que contribuam para o valor agregado.	(Demchenko, Grosso, De Laat, & Membrey, 2013)
	Capacidade de transformar os dados em valor comercial.	(Budhathoki, Dasgupta, & Jain, 2018)

Fonte: Elaboração própria.

No mundo atual de milhares de dados, não cabe mais analisar o passado e obter intuições sobre o futuro de forma manual. “É impossível realizar essa tarefa em banco de dados tradicional. Uma empresa não quer mais saber daqui a um mês o que aconteceu há 40 dias” (Florissi, 2013). *Big data* transformará a forma como vivemos, trabalhamos e pensamos. Em virtude disso, é considerado uma revolução e análise é seu núcleo principal. A análise de dados envolve várias abordagens, tecnologias e ferramentas (Jagadish, et al., 2014).

Machine learning é um dos componentes fundamentais da análise de dados. Este será um dos principais impulsionadores da revolução do *big data* (Manyika, et al., 2011). A razão para isso é sua capacidade de aprender e fornecer intuições, decisões e previsões baseados em dados (Rouse, 2018). O processo de *machine learning* é baseado em estatísticas e, similarmente à análise estatística, pode extrair tendências de dados. Com isso chega-se ao ponto de tratar do principal conceito deste item que é *machine learning*.

2.2.3. Conceito de *Machine Learning*

Neste ponto, surge o seguinte questionamento: qual a diferença entre inteligência artificial e *machine learning*? Então aproveita-se o momento e se expande para diferenciá-los de *deep learning* (aprendizado profundo).

A inteligência artificial é o conceito mais amplo dos três em questão, que consiste em tudo. John McCarthy, amplamente reconhecido como um dos padrinhos da IA, definiu inteligência artificial como: “a ciência e a engenharia de fabricação de máquinas inteligentes” (McCarthy, 2007). Sempre que uma máquina conclui tarefas com base em um conjunto de regras estipuladas para resolver problemas (algoritmos), esse comportamento “inteligente” é chamado de inteligência artificial (Luger G. , 2009).

O segundo conceito é o de *machine learning*, que em 1959, com Arthur Samuel, um dos pioneiros, definiu como um “campo de estudo que dá aos computadores a capacidade de aprender sem serem programados explicitamente” (Samuel, 1959). “É um conjunto de regras e procedimentos, que permite que os computadores possam agir e tomar decisões baseadas em dados ao invés de ser explicitamente programados para realizar uma determinada tarefa. Programas de *machine learning* também são projetados para aprender e melhorar ao longo do tempo quando expostos a novos dados” (Matos, 2015).

A Figura 2 resume o cenário que se encontram estes três componentes:

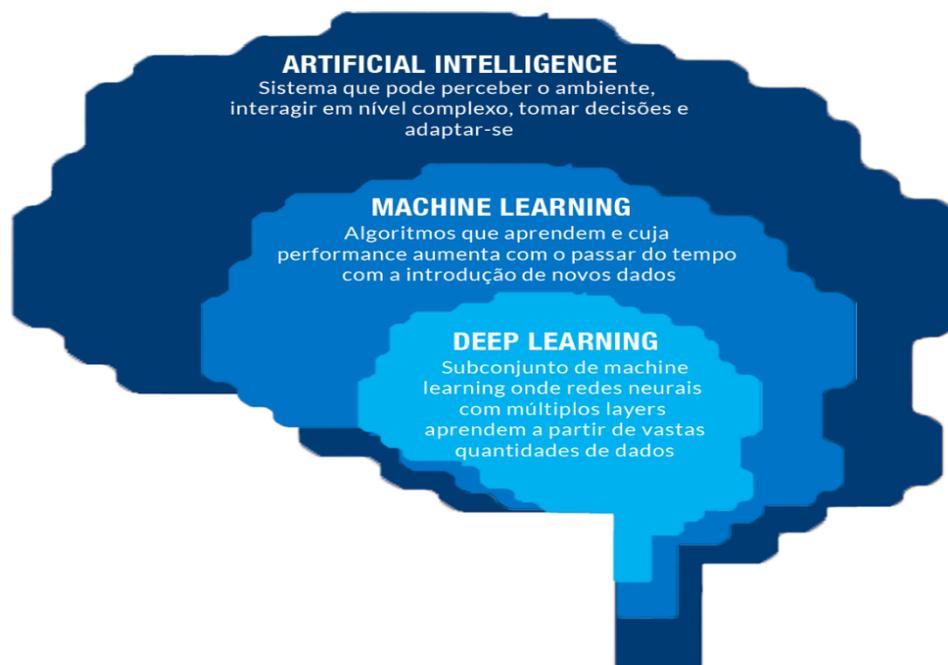


Figura 2 – Machine learning e deep learning no contexto da inteligência artificial.

Fonte: (Nascimento, 2017).

Já em consulta a enciclopédia Britannica pode-se abstrair que “*machine learning* é um subcampo da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial” (Enciclopédia Britannica, 2019).

É uma tecnologia altamente interdisciplinar que se baseia em ideias de muitos tipos de campos, tais como inteligência artificial, teoria da otimização, teoria da informação, estatística, ciência cognitiva, controle ótimo e muitas outras disciplinas da ciência, engenharia e matemática (Mitchell, 1997). Por abranger diversas áreas de domínio científico, o aprendizado de máquina traz grande impacto sobre a ciência e a sociedade.

Em outras palavras, são máquinas que aprendem através dos dados resultantes dos algoritmos sem nenhuma programação prévia e que se otimizam ao longo do tempo por seu constante aprendizado. São sistemas que: “pensam como um humano; agem como um humano; pensam racionalmente ou agem racionalmente” (Russell & Norvig, 2013).

Então, em resposta a primeira parte do questionamento, pode-se afirmar que *machine learning* está relacionado ao aprender, enquanto inteligência artificial trata de vários outros aspectos da inteligência humana.

Quanto a *deep learning*, um subconjunto do *machine learning*. Já *machine learning* é um subconjunto da inteligência artificial. Em outras palavras, todo *machine learning* é inteligência artificial, mas nem toda inteligência artificial é *machine learning*, e assim por diante. *Deep learning* seria a próxima evolução de *machine learning* que também é chamado de aprendizado estruturado profundo, aprendizado hierárquico ou aprendizado de máquina profunda. Surge do princípio básico de que existem várias etapas de análises diferentes, onde cada etapa recebe como entrada o resultado da etapa anterior. Criando assim uma ideia de profundidade no aprendizado (Schade, 2018).

Então conclui-se este questionamento dizendo que *deep learning* pode descobrir automaticamente os recursos a serem usados para classificação, enquanto que *machine learning* exige que esses recursos sejam fornecidos manualmente.

2.2.4. Categorias de *Machine Learning*

Normalmente, tipificam-se *machine learning* em três categorias conforme a Tabela 3. Estas categorias de ML variam de acordo com o nível de intervenção humana na rotulação dos dados. Algoritmos de aprendizado supervisionado constroem um modelo matemático de um conjunto de dados que contém tanto as entradas quanto as saídas desejadas (Russell & Norvig, 2013). Tem como objetivo a aprendizagem de um padrão para mapear as entradas para a saída. Dentre as técnicas mais conhecidas para resolver problemas de aprendizado supervisionado estão máquina de suporte vetorial, Naïve Bayes, Modelo de Markov, redes Bayesianas e redes neurais. *Machine learning* supervisionado é a área que concentra a maioria das aplicações de sucesso e onde os problemas já estão bem definidos.

Na aprendizagem não supervisionada, não há variável ou resultado para prever ou estimar (Turkson, Baagyere, & Wenya, 2016). Os sistemas não possuem rótulos e nem sabe a resposta certa. O objetivo aqui é explorar os dados e encontrar uma estrutura em entradas. As principais técnicas são: K-means, modelo de mistura Gaussiana, modelo de mistura de processo Dirichlet e X-means.

Tabela 3 – Categorias de machine learning

Categorias de ML	Tarefa	Norma de distinção	Algoritmo de ML
Aprendizado supervisionado	Classificação Regressão Estimação	Classificadores Computacionais	Máquina de suporte vetorial
		Classificadores estatísticos	Naïves Bayes
			Modelo oculto de Markov
			Redes Bayesiana
Classificadores cognitivos	Redes Neurais Artificiais		
Aprendizado sem supervisão	Clustering Predição	Paramétricos	K-means
			Modelo de mistura Gaussiana
		Não-paramétricos	Modelo de mistura de processo Dirichlet
			X-means
Aprendizado por reforço	Tomada de decisão	Modelo <i>free</i>	Q-learning
			R-learning
		Modelo <i>based</i>	TD-learning
			Aprendizado Sarsa

Fonte: (Qiu, et al., 2016)

Aprendizado de reforço é o *machine learning* que permite que *softwares* e máquinas identifiquem automaticamente o comportamento ideal dentro de um quadro preciso, de modo a maximizar seu desempenho (Giryes & Elad, 2011). O sistema interage com um ambiente dinâmico e deve realizar uma determinada meta sem orientação ou professor. Devido à sua generalidade, o campo é estudado em muitas outras disciplinas, como teoria dos jogos, teoria de controle, pesquisa operacional, teoria da informação, otimização baseada em simulação, sistemas multiagentes, inteligência de enxames, estatística e algoritmos genéticos (Bertsekas D., 2012). As técnicas mais utilizadas em problemas de aprendizado por reforço são *q-learning*, *r-learning*, *td-learning* e aprendizado sarsa. Por fim cabe ainda relatar que as aplicações de sucesso que mais se destacam no aprendizado por reforço são as de veículos autônomos e as de jogos contra um oponente humano.

A seguir passou por descrever sobre os principais algoritmos de *machine learning*.

2.2.5. Algoritmos de *Machine Learning*

A. Máquinas de Suporte Vetorial

As máquinas de suporte vetorial (SVM), também conhecidas como redes de vetores de suporte, são técnicas de *machine learning* supervisionado, agrupadas para serem empregadas na classificação e regressão (Cortes & Vapnik, 1995).

As SVM é um dos algoritmos de *machine learning* mais utilizado e tem obtido sucesso na resolução de diversos problemas. A boa propriedade de generalização de um SVM depende de um subconjunto chamado *Support Vectors* (SV). Estes que são vetores que descrevem o limite de decisão. Infelizmente, o treino de SVM tradicional é muito lento e consome um espaço razoável. Então, é possível que não seja uma boa opção para lidar com problemas de larga escala (Zheng, et al., 2012).

B. Naïve Bayes

Essa técnica de classificação é baseada no teorema de Bayes, representa um método estatístico e de aprendizado supervisionado, que pressupõe independência entre os preditivos, portanto, a presença de uma característica particular numa classe não obriga a sua existência em outra (Ray, 2017; Flores, et al., 2018).

Como pontos positivos de Naïve Bayes pode-se citar que calcula as probabilidades explícitas para hipóteses, é fácil de construir e particularmente útil para conjuntos de dados muito grandes. O seu ponto negativo está relacionado a hipótese de preditivos independentes. É quase impossível obtermos um conjunto de preditivos completamente independentes (Ray, 2017).

C. Modelo Oculto de Markov

Modelo oculto de Markov, ou HMM, é um modelo estatístico no qual o sistema que está a ser modelado é considerado um processo de Markov com estados não observados (isto é, oculto). O adjetivo oculto se refere à sequência de estados pela qual o modelo passa, não aos parâmetros do modelo; o modelo ainda é chamado de modelo oculto de Markov, mesmo que esses parâmetros sejam conhecidos com exatidão. O HMM pode ser representado como a rede bayesiana dinâmica mais simples. A matemática por trás do HMM foi concebida por L. E. Baum e colaboradores (Baum, 1972).

O HMM está relacionado ao problema ótimo de filtragem não linear de Ruslan L. Stratonovich, que foi o primeiro a descrever o procedimento de avanço / recuo. (Stratonovich, 1960)

Os modelos ocultos de Markov são especialmente conhecidos por sua aplicação no aprendizado por reforço e reconhecimento de padrões temporais, como fala, caligrafia, reconhecimento de gestos, marcação de parte da fala, sequência de partitura musical, descargas parciais e bioinformática (X. Huang, Acero, & Hon, 2001; Pardo & Birmingham, 2005; Satish L, 2003).

D. Redes Bayesianas

Uma rede bayesiana, ou rede de crenças, ou modelo gráfico acíclico direcionado (GAD), é um modelo probabilístico que representa um conjunto de variáveis aleatórias e a sua independência condicional com um gráfico acíclico direcionado. Pode ser usada tanto para aprendizado supervisionado quanto para aprendizado não supervisionado.

Em geral, existem duas abordagens para aprender a estrutura da rede bayesianas: uma é baseada em pontuação e a outra em restrições. A abordagem baseada em pontuações usa uma função de pontuação para medir quão bem uma estrutura se ajusta aos dados de treino e tenta encontrar a estrutura que tem a melhor pontuação. Já abordagem baseada em restrições usa testes de independência para determinar as arestas e as direções.

A classificação da rede bayesiana é uma técnica muito eficaz para modelagem preditiva e também para seleção de variáveis. Muitos pacotes de software possuem implementações dessas diferentes estruturas. Uma medida fundamental para a avaliação de diferentes pacotes é o paralelismo e a capacidade de lidar com dados amplos. Tal como enfatiza Dean, os classificadores de rede bayesianas podem executar bem em dados que são centenas de milhares de colunas e, dependendo da implementação, até trabalhar em dados densos (Dean, 2014).

E. Redes Neurais Artificiais

É um modelo baseado nas redes neurais biológicas, onde nós conectados entre si, formam sua estrutura. São da categoria de aprendizagem supervisionada. Se utilizam de métodos estatísticos para calcular ou aproximar funções que dependem de um grande número de entradas. As RNA são aplicadas com maior frequência quando o volume de entrada é imenso para as abordagens convencionais de aprendizagem automática.

Em outras palavras, são procedimentos computadorizados criados para elaborar modelos de estrutura neural de organismos inteligentes e aprenderem com os resultados anteriores (Sriram, 1997). As redes neurais suportam algoritmos de classificação e regressão. São estruturas criadas para trabalharem conforme o cérebro humano. Para terem performance optam por uma grande quantidade de unidades de processamento simples, denominados neurónios (Haykin, 2001).

Uma das propriedades mais importantes das RNA é a capacidade de simular o aprendizado, isto é, usar novos dados para ajustar o modelo e melhorar o desempenho. Aprendizagem é um processo pelo qual os parâmetros livres da RNA são ajustados, através de um processo de estimulação, do ambiente em que ela está inserida. Dependendo da forma como ocorre a modificação dos parâmetros, determina-se qual é a estratégia de aprendizagem (Haykin, 2001).

As redes neurais artificiais apresentam como vantagens a sua capacidade de aprender através de uma fase chamada de aprendizagem. A sua auto-organização de informação que pode ser disponibilizada ao utilizador. É bastante tolerante as falhas, tem a flexibilidade de lidar com grandes mudanças na informação de entrada, e a sua estrutura paralela permite a obtenção de respostas em tempo real.

Atualmente as redes neurais estão a ser usadas numa ampla variedade de campos para modelar tudo, desde fraudes com cartões de crédito, perfurações de poço de petróleo, escolha de ações para investimento, como identificação de faces em fotos, se uma pessoa é suscetível de responder a um anúncio, se uma pessoa é suscetível de contrair uma doença, e se um medicamento é suscetível de curá-lo (Dean, 2014).

F. Algoritmo K-means

Antes de tratar deste algoritmo é importante introduzirmos o conceito de *clustering*. *Clustering* é o processo de agrupar grandes conjuntos de dados em diferentes *clusters* independentes, de modo que os objetos de dados em um *cluster* sejam semelhantes entre si e diferentes dos objetos de dados presentes em outros *clusters*. O armazenamento em *cluster* ajuda a distinguir os dados criando *clusters* para que as operações necessárias possam ser executadas apenas nos dados presentes em um *cluster* desejado, em vez do conjunto de dados completo (Asha, Patyal, & Chen, 2018).

K-means é um algoritmo de *clustering* do tipo *machine* não supervisionado, que é usado quando se tem dados sem rótulo (ou seja, dados sem categorias ou grupos definidos). O objetivo desse algoritmo é encontrar grupos nos dados, com o número de grupos representados pela variável k . O algoritmo trabalha iterativamente para atribuir cada ponto de dados a um dos grupos k com base nos recursos fornecidos (Trevino, 2016).

É uma das técnicas de *clustering* mais usadas por causa da sua simplicidade e velocidade. Além disso, pode produzir bons clusters através do particionamento dos dados em k clusters assinando cada objeto com o seu centróide de cluster mais próximo (o valor médio das variáveis para todos os objetos nesse cluster específico) com base na medida de distância usada. É mais robusto para diferentes categorias de variáveis e também mantém o desempenho para grandes conjuntos de dados, que são comuns na segmentação (Dean, 2014).

O algoritmo *k-means* sempre converge para o mínimo local. Antes da convergência, cálculos de distância e centros de clusters são feitos enquanto *loops* são executados um número de vezes, onde o inteiro positivo t é conhecido como o número de iterações *k-means*. O valor preciso de t varia a depender dos centros iniciais do cluster. A distribuição dos pontos de dados tem um relacionamento com o novo centro de armazenamento em cluster. Portanto, a complexidade do tempo computacional do algoritmo *k-means* é nkt onde, n é o número de todos os objetos de dados, k é o número de clusters, t é a iteração do algoritmo. Geralmente requerendo $k \ll n$; e $t \ll n$.

Assim, pode-se dizer que uma deficiência deste algoritmo é que ele calcula a distância de cada objeto de dados para cada centro de cluster em cada iteração. No entanto, por experimentos, descobriu-se que não é necessário calcular essa distância a cada vez. Assim, no decorrer de várias iterações ocupam-se um longo tempo de execução, afetando a eficiência do *clustering* (Na, Xumin, & Yong, 2010).

Este é um algoritmo versátil que pode ser usado para qualquer agrupamento (Trevino, 2016):

- Segmentação comportamental;
- Segmento por histórico de compras;
- Segmento por atividades no aplicativo, site ou plataforma;
- Definir pessoas com base em interesses;

- Criar perfis com base no monitoramento de atividades;
- Categorização de inventário;
- Inventário do grupo por atividade de vendas;
- Inventário do grupo por métricas de fabricação;
- Medidas para classificação de sensores;
- Detetar tipos de atividade em sensores de movimento;
- Agrupar imagens e áudios separados;
- Identificar grupos no monitoramento da saúde;
- Detetar anomalias;
- Agrupar atividade válida para limpar a detenção de *outliers*.

G. Modelo de mistura Gaussiana

Algoritmos de mistura gaussiana são modelos probabilísticos para representar subpopulações normalmente distribuídas dentro de uma população total. Este afirma que todos os pontos de dados gerados são derivados de uma mistura de distribuições gaussianas finitas que não possuem parâmetros conhecidos (Santosh, et al., 2013).

Os modelos de misturas, em geral, não exigem saber a qual subpopulação que pertence, permitindo que o modelo aprenda as subpopulações automaticamente. Como a designação de subpopulação não é conhecida, isso constitui uma forma de *machine learning* não supervisionado. (McGonagle, Pilling, & Tembo, 2019)

Os parâmetros dos modelos de mistura gaussiana são derivados da estimação máxima *a posteriori*, ou de um algoritmo de maximização de expectativa iterativa de um modelo anterior que foi bem treinado.

Matematicamente, os modelos de mistura gaussiana são um exemplo de uma função de densidade de probabilidade paramétrica, que pode ser representada como uma soma ponderada de todas as densidades de componentes gaussianos. Seu modelo consiste em matrizes de covariância, pesos de mistura e vetores médios de cada densidade de componente presente. Os gaussianos são capazes de modelar as correlações dos elementos do vetor de recursos graças à combinação linear da base de covariância diagonal (Techopedia, 2019).

Um ponto importante a ser mencionado é que os resultados de agrupamentos GMM são altamente sensíveis à inicialização de centros e misturas, porque só garantem encontrar uma solução ótima local em relação à inicialização. Para obtenção de centros iniciais razoáveis é indicado o uso do algoritmo KKZ em cima dos nós da subamostra encontrados do conjunto de dados de treino. A grande vantagem de se fazer amostragem por KKZ é que os *outliers* da distribuição de dados reais, é pouco possível de serem selecionados, então os centros iniciais estarão mais perto da distribuição de dados de entrada, mas mesmo assim suficientemente separado para serem identificados como centros (Katsavounidis, Kuo, & Zhang, 1994; (J. He, et al., 2004).

Os modelos de mistura gaussiana são muito utilizados para o tratamento de modelos de dados, principalmente quando os dados são oriundos de diversos grupos. Além disso, são usados em sistemas biométricos, onde o modelo paramétrico ajuda a entender as características ou medidas relacionadas a dados de fala e também foram aplicados extensivamente no rastreamento de objetos em detecção de eventos e identificação humana em videovigilância.

Outro bom exemplo está relacionado a retornos financeiros que geralmente se comportam de maneira diferente em situações normais e durante períodos de crise. Às vezes neste caso, o modelo usado é um de difusão em saltos ou uma mistura de duas distribuições normais. Por este motivo, um modelo de mistura para dados de retorno parece bem razoável (Dinov, 2008).

H. Modelo de Mistura de Processo Dirichlet

É um modelo bayesiano, não paramétrico, pertencente a uma classe de modelos de mistura infinita. Não se impõe nenhum conhecimento prévio sobre o número de *clusters* (K). Os modelos DPMM aprendem a quantidade de *clusters* dos dados usando um prior não paramétrico baseado no Dirichlet process (DP) antes de explorar modelos de mistura. A seleção automática de modelos leva à economia computacional da validação cruzada do modelo para múltiplos valores de K (Jinsong, Shuang, Diyin, & Hao, 2017) (Smolyakov, 2017).

O DPMM descobre automaticamente diferentes padrões com o tempo. Aprende os parâmetros iniciais para cada padrão simultaneamente. Desta forma, pode encurtar o

tempo de convergência do modelo para o sistema real responder com uma boa seleção de parâmetros iniciais.

O modelo de mistura de processo de Dirichlet é uma abordagem bem conhecida e tornou-se popular devido à sua natureza não supervisionada e não paramétrica (Santhosh, Dogra, & Roy, 2017).

Possui bons resultados em processo de *clustering* de objetos em movimento usando fluxo ótico. Ele também tem muitas aplicações em intercetação e rastreamento, análise de tráfego, análises de imagens médicas e aprendizado por reforço (Jinsong, et al., 2017).

I. *Q-learning*

O *Q-learning* é uma técnica ML de aprendizado por reforço. O objetivo deste é aprender uma política, que diz a um agente qual ação tomar e sob quais circunstâncias. Não requer um modelo do ambiente e pode lidar com problemas com transições e recompensas estocásticas, sem exigir adaptações. Em outras palavras, *Q-learning* baseia-se num valor que é usado para encontrar a política de seleção de ação ideal usando uma função Q.

Como em qualquer processo finito de decisão de Markov (FMDP), o algoritmo ML de *Q-learning* encontra uma política que é ótima no sentido que maximiza o valor esperado da recompensa total sobre todos os passos sucessivos, a partir do estado atual (Melo, 2007). O *Q-learning* pode identificar uma política ótima de seleção de ação para qualquer FMDP, dado o tempo de exploração infinito e uma política parcialmente aleatória (Melo, 2007). “Q” nomeia a função que retorna a recompensa usada para fornecer o reforço e preserva a qualidade de uma ação tomada num determinado estado (Tambet, 2015).

Q-learning é insensível à exploração, isto é, os valores Q convergem para os valores ótimos, independente de como o agente se comporta enquanto os dados estão a ser coletados. Isso significa que, embora a questão “*exploration-exploitation*” deva ser abordada no *Q-learning*, os detalhes da estratégia de exploração não afetarão a convergência do algoritmo de aprendizado. Por estas razões, o *Q-learning* é o mais popular e parece ser o algoritmo mais eficaz em *machine learning* com o reforço atrasado (Leslie, Littman, & Moore, 1996).

O sistema de aprendizado “reforço secundário”, precursor do *Q-learning* (Sutton & Barto, 2017). Tem como princípio modelar valores de estado via retro propagação, ou seja, o valor de estado "v" é retro propagado para a situação anteriormente encontrada.

Em 2015, a Google patenteou um aplicativo de *Q-learning* para *deep learning*, intitulado “*deep reinforcement learning*” ou “*deep Q-learning*” que pode reproduzir jogos do Atari 2600 em níveis humanos especializados (US Patent, 2015).

Por fim elenca-se alguns dos usos do algoritmo *Q-learning*: gestão de recursos em clusters de computadores, controle de semáforo de trânsito, robótica, configuração de sistemas WEB, otimização de reações químicas, jogos (Garychl, 2018).

J. *R-learning*

Assim como o *Q-learning*, o *R-learning* de Schwartz também é uma técnica de aprendizado de reforço baseado em valor de recompensa (Schwartz, 1993). É conhecido pelo caso contínuo não descontado, ou maximização da “recompensa média por etapa de tempo”, ou ainda, “caso de recompensa média”. Os principais mentores do aprendizado por reforço para tarefas contínuas não descontadas: (Schwartz, 1993; Tadepalli & Ok, 1994; Mahadevan, 1996; Bertsekas & Tsitsiklis, 1996). Além desses tiveram (Nordstrom & Carlstrom, 1997) que sugeriram o exemplo de enfileiramento de controlo de acesso (Sutton & Barto, 2015).

A principal diferença entre os dois é que o *Q-learning* de Watkins é um método para otimizar recompensas descontadas (acumuladas), fazendo com que recompensas futuras sejam menos priorizadas do que as de curto prazo (Watkins, 1989). Já o *R-learning* otimiza a recompensa média, pesando tanto o futuro distante quanto as recompensas em curto prazo (Schwartz, 1993).

Embora otimizar a recompensa média pareça mais natural, ela também apresenta alguns problemas, como tornar o agente “preguiçoso”. Houve algumas tentativas de contornar isso, mas o *Q-learning* continua a ser o algoritmo mais popular. Além disso, comprovou-se que o *Q-learning* converge para a política ótima no caso tabular, diferentemente do *R-learning* (Everitt, 2017).

K. TD-learning

Sutton e Barto, divide aprendizado de reforço em três classes: programação dinâmica, métodos de Monte Carlo e aprendizagem por diferença temporal, dos quais este último é chamado a ideia mais central e nova no aprendizado por reforço. A programação dinâmica é baseada na Equação de Bellman que divide um problema em problemas menores. Dividindo uma grande tarefa em etapas menores, essa abordagem depende de um modelo perfeito do ambiente. Os métodos de Monte Carlo não precisam de um modelo perfeito do ambiente de aprendizagem. A partir da experiência em forma de sequências de amostras de recompensa (ação-estado) elas podem se aproximar das recompensas futuras. No entanto, os métodos só são atualizados após uma sequência completa, quando o estado final é atingido (Sutton & Barto, 2015).

Os métodos de diferença temporal combinam ambos os procedimentos. Enquanto os métodos de Monte Carlo apenas ajustam as suas estimativas uma vez que o resultado é conhecido, as técnicas de TD adequam as previsões para que correspondam a previsões mais precisas sobre o futuro antes que o resultado seja conhecido (Sutton R., 1988).

Em outras palavras, a aprendizagem de diferença temporal (TD) é uma técnica não supervisionada de aprendizagem por reforço. Esta é treinada por *bootstrapping*, a partir da estimativa atual da função de valor (Sutton & Barto, 2015; McClelland, 2015).

Os métodos de diferença temporal estão relacionados a aprendizagem de animais. Não possuem conexão inerente às arquiteturas de redes neurais. O aprendizado de TD resolve o problema da atribuição de crédito temporal, ou seja, a questão de atribuir a culpa pelo erro pela sequência de previsões feita pelo agente de aprendizado (McClelland, 2015).

Dentre alguns estudos importantes relacionados a este algoritmo, estão (Tsitsiklis & Van Roy, 1996) com abordagens de Cadeias de Markov descontinuadas. No contexto das cadeias de Markov absorventes tem: (Bertsekas & Tsitsiklis, 1996). Para além desses, menciona-se (Bradtke & Barto, 1996) que evoluiu o TD para LSTD, (Boyan, 2002) que apresentou uma versão estendida do LSTD e (Lagoudakis & Parr, 2003) que demonstrou LSTDQ como o algoritmo que aprende uma aproximação para a função de valor de ação. Onde o algoritmo TD (λ) aproxima a função de valor de estado. Para o cálculo de trajetórias de amostra $V \pi$ de uma política, quando dado. O algoritmo LSTDQ recebe uma fonte de conjuntos de amostras da forma (s, a, r, s_0) a

partir da qual escolhe amostras de treino arbitrariamente. Por esse método, a função de valor de ação de diferentes políticas pode ser aproximada a usar a mesma fonte de amostras. Contudo, exige-se uma representação linear da função valor.

Ao final cabe-se mencionar que esse método é eficaz para tarefas que possuem um espaço de estado enumerável. No entanto, em muitas tarefas, e certamente naquelas com um espaço de estado contínuo, não pode esperar enumerar todos os estados possíveis. Mesmo que conseguisse, há muito para uma implementação na prática. A propagação reversa resolve o problema da atribuição de crédito estrutural. Ou seja, o problema de como ajustar os pesos na rede para minimizar o erro. Um dos maiores benefícios das redes neurais é, naturalmente, a sua capacidade de generalizar o aprendizado em estados semelhantes. Assim, combinar tanto TD quanto propagação reversa resulta num agente que pode aprender de forma flexível a maximizar a recompensa em múltiplas etapas de tempo e também descobrir semelhanças estruturais em padrões de entrada que lhe permitem generalizar as suas previsões sobre novos estados. Certamente, este é um acoplamento poderoso (McClelland, 2015).

2.3. *Machine Learning* para Risco de Crédito

Nas secções anteriores foram abordadas as partes mais conceituais dos dois pilares deste estudo: *machine learning* e risco de crédito. Nesta secção faz-se a combinação destes dois temas. Apresenta as principais vantagens e aplicações que *machine learning* pode proporcionar aos processos de risco de crédito. Elabora um estudo bibliométrico que demonstra a contribuição do conhecimento científico derivado das publicações, elencando os principais autores e obras da relação destes assuntos.

2.3.1. Vantagens

Para um ser humano, monitorar todas as variáveis envolvidas numa gestão de risco de crédito e ainda ter um resultado realmente assertivo é um fato difícil de se alcançar e que requer muito tempo. Esta é a razão principal que *machine learning* pode impulsionar economias e aumentar o desempenho de crédito nas instituições bancárias de todo o mundo. *Machine learning* melhora a análise de classificação do cliente, torna a predição mais assertiva, consegue extrair informações estruturadas de dados não-estruturados, otimiza os parâmetros de regressão, diminui o erro de aproximação e calcula o melhor resultado.

As Tabelas de 4 a 6 detalham melhor estas vantagens:

Tabela 4 – Vantagens de ML no processo de aprovação de crédito

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Naïves e redes bayesianas	Predição	Previsão de <i>default</i> quando os dados estão balanceados	(Chung , Ho, & Hsu , 2011) (Brown & Mues, 2012) (Marqués, García , & Sánchez, 2012) (Crone & Finlay, 2012) (G. Wang, et al., 2012) (Pascual, Martínez, & Alamillos, 2014) (Bijak & Thomas, 2012) (Bücher, Kampen, & Krämer, 2013) (Nikolic, et al., 2013) (Louzada-Neto & Ara, 2012) (Louzada-Neto, Ferreira-Silva , & Diniz, 2012)
Rede Neurais	Classificação <i>Credit score</i>	Eficiência e eficácia no processo de classificação de bom e maus pagadores. ³	(Tam & Kiang, 1992 citado por Forti, 2018); (Hens e Tiwari, 2012 citado por Silva, 2014); (Blanco, et al., 2013); (De-La-Veja, et al., 2013); (Harris, 2013); (Kruppa, et al., 2013); (Brown & Mues, 2012); (Louzada-Neto & Ara, 2012) (Marqués, García , & Sánchez, 2012); (Crook, Edelman, & Thomas, 2007); (Steiner, et al., 2007);
	Predição	Melhor abordagem às previsões de séries financeiras, análise de crédito, previsão de insolvência de empresas.	(Perera, 2013)

³ Redes neurais são melhores na identificação de maus pagadores do que bons pagadores.

Tabela 4 – Continuação.

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Regressão logística	Classificação <i>Credit score</i>	Possibilidade e facilidade de interpretação e sua ótima estabilidade ao longo do tempo. ⁴	(Forti, 2018)
		Eficiência e eficácia no processo de classificação de bom e maus pagadores. ⁵	(Feldman & Gross, 2005) (Brown & Mues, 2012) (Marqués, García , & Sánchez, 2012) (Bijak & Thomas, 2012) (Crone & Finlay, 2012) (G. Wang, et al., 2012) (Pascual, Martínez, & Alamillos, 2014) (Brito & Assaf Neto, 2008) (Minussi, Damacena, & Ness Jr., 2002) (Dantas & Souza , 2008) (Louzada-Neto & Ara, 2012)
Árvore de decisão	Análise	Melhora a abordagem na determinação regras de concessão de crédito.	(Perera, 2013) (Brown & Mues, 2012) (Crone & Finlay, 2012) (Kim & Sohn, 2012) (G. Wang, et al., 2012)
Métodos <i>Ensemble</i> (<i>boosting e random forest</i>)	Predição	Previsão de <i>default</i> quando os dados não estão balanceados	(Brown & Mues, 2012)
		Maior precisão do que RL, contradizendo Melissa Forti.	(Florez-Lopez & Ramon-Jeronimo, 2015)
	Classificação <i>Credit score</i>	Melhor ordenação nos scores.	(Forti, 2018)
	Análise	Análise de crédito de forma rápida e de fácil interpretação.	(Abellán & Mantas, 2014) (Kruppa, et al., 2013);

⁴ Regressão logística é o algoritmo de *machine learning* mais utilizado no mercado.

⁵ Regressão logística é melhor na identificação de bons pagadores do que maus pagadores.

Tabela 4 – Continuação.

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Análise discriminante	Segmentação	Melhor categorização de risco de crédito de uma pessoa.	(Elliott, Siu, & Fung, 2014) (Brown & Mues, 2012) (Crone & Finlay, 2012) (Louzada-Neto & Ara, 2012) (Marqués, García, & Sánchez, 2012) (A. Guimarães & Moreira, 2008) (Hair Jr, et al., 2007) (I. A. Guimarães & Chaves Neto, 2002)
SVM	Predição	Melhora a avaliação do risco de crédito. Previsão de <i>default</i> (incumprimento).	(R. Silva, Ribeiro, & Matias, 2016) (Bellotti & Crook, 2013) (Perera, 2013) (Z. Huang et al., 2004)
Modelos híbridos	Classificação <i>Credit score</i>	Eficiência e eficácia no processo de classificação	(Doumpos, Zopounidis, & Adriosopoulos, 2014)

Fonte: Elaboração própria.

Tabela 5 – Vantagens de ML no processo de acompanhamento de crédito

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Redes Bayesianas	Classificação <i>Behavior score</i>	Demonstrar transparência às relações de causa e efeito do crédito e a volatilidade dos diversos tipos de lucro do banco	(Camarinha, 2009)
	Auditoria	Eficiência e eficácia aos processos de auditoria interna de crédito.	(Laws & A. O'Hagan, 2003) (Sarkar & Sriram, 2001)
Náive Bayes	Risco de Corrupção	Identificação de risco de corrupção nas operações do cliente e funcionário.	(Balaniuk, 2013)
Regressão logística	Auditoria	Melhor aceitação pelos órgãos reguladores devido a esta ter um melhor controle e rastreabilidade do modelo.	(Forti, 2018)
	Deteção de Fraude	Identificação de relatórios financeiros falsificados a partir de dados disponibilizados publicamente.	(Spathis, 2002 citado por Kadam & Raval, 2014)

Tabela 5 – Continuação.

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Redes Neurais	Deteção de Fraude	Maior assertividade na identificação de uma operação ser fraudulenta.	(Coelho, Raittz, & Trezub, 2006) (Aleskerov, Freisleben & Rao; Hassibi; Dorronsoro, et al.; Syeda, Zhang & Pan citado por Yufeng, et, 2004) (Ghosh & Reilly; Brause, Langsdorf & Hepp citado por Bolton & Hand, 2002)
	Deteção de Crime	Identificação de crime de lavagem de dinheiro	(Chartier e Spillane citado por Bolton & Hand, 2002)
Árvore de decisão	Predição	Melhor previsibilidade de continuidade de negócios com o cliente. (<i>Going Concern Prediction</i>).	(Koh & Low citado por Kadam & Raval, 2014)
<i>Random Forest</i>	Deteção de Fraude	Melhor performance na identificação de fraudes.	(Bhattacharyya, 2011)
Modelos Híbridos	Risco de Corrupção	Identificação de risco de corrupção nas operações do cliente e funcionário.	(R. S. Carvalho et al., 2014)

Fonte: Elaboração própria.

Tabela 6 – Vantagens de ML no processo de recuperação de crédito

<i>Machine Learning</i>	<i>Credit Risk</i>	<i>Vantagens</i>	<i>Referências</i>
Regressão logística	Classificação <i>Collection Score</i>	Possibilidade e facilidade de interpretação e sua ótima estabilidade ao longo do tempo.	(Forti, 2018)
Análise de Sobrevivência	Classificação <i>Collection Score</i>	Desenvolvimento de <i>Collection score</i> em termos estratégicos.	(Machado, 2016) (Souza, 2000)
Modelos híbridos	Classificação <i>Collection Score</i>	Melhor assertividade na classificação de rating de crédito	(Y. -S. Chen & Cheng, 2013)

Fonte: Elaboração própria.

Logo se percebe que *machine learning* proporciona diversos modelos, vantagens e aplicações. Não existe pelos autores, uma unanimidade entre os modelos. A unanimidade está em dizer que cada problema, tem a sua peculiaridade que encaixa num determinado modelo. Cabe a instituição bancária eleger, através do seu perfil, estratégias e testes o seu

modelo campeão. Outra alternativa, seria seguir estudos mais recentes de utilização de modelos híbridos (Y. -S. Chen & Cheng, 2013).

Independente da escolha de modelo, o que deve ser explicado é que *machine learning*:

- Define uma probabilidade de incumprimento a muito curto prazo (IDC, 2019);
- Encontra padrões de comportamento não supervisionados (IDC, 2019);
- Proporciona uma avaliação mais profunda das variáveis de crédito que ajuda na inclusão de variáveis que sejam importantes e exclusão das variáveis que não se mostrem relevantes, com vantagens de proporcionar modelos de gestão de crédito mais, sucintos, precisos, flexíveis e com capacidade de otimização (Duan & Shrestha, 2011). Minimizando esforço computacional (Hens & Tiwari, 2012) citado por (R. A. Silva, 2014), economia de tempo na execução e melhor acurácia nas decisões.
- Oferece análises de risco de crédito eficientes, minimizam o risco de falhas (Lu et al., 2013 citado por R. A. Silva, 2014; Lempriere, 2017).

No âmbito operacional, *machine learning* simplifica as operações de crédito, aumenta a segurança da informação e garante a conformidade regulamentar dos contratos de crédito, evitando comportamentos criminosos de funcionários e clientes (corrupção e lavagem de dinheiro). Melhora a decisão de negócios, uma vez que podem de forma ágil analisar taxas de incumprimento do passado, dados geográficos, informações de diversas fontes do cliente prever possíveis resultados futuros. (Lempriere, 2017)

Além disso, *machine learning* oferece ao risco de crédito, uma abordagem alternativa de digitalização de processos centrais, onde os próprios algoritmos determinam os condutores de risco. Minimizando assim, intervenções manuais ao fazer modelagem, simplificação, padronização e automação. Permitindo a análise de um volume maior de informações do cliente a custos mais baixo de operação. Melhora a qualidade geral de dados, recursos de agregação, e relatório de oportunidades. Podendo inovar na disponibilização de relatórios, substituindo o papel por uma solução interativa que oferecesse uma grande quantidade de informações do risco de crédito em tempo real. Tais relatórios iriam permitir aos utilizadores fazer a sua própria análise, e aos bancos tomar decisões mais rapidamente. Além disso, relatórios de risco de crédito em tempo real e com base em dados transparentes e coerentes, iriam melhorar a qualidade das decisões baseadas em fatos, ajudar os bancos a identificar riscos potenciais ainda mais cedo, e

permitir uma resposta mais oportuna. *Machine learning* pode alinhar a gestão de crédito com a estratégia financeira da empresa, otimizando o balanço. Indicando o melhor momento de aumentar ou diminuir uma carteira de empréstimos (Härle, Havas, & Samandari, 2016).

Quando integrado com sistemas de *big data* podem aprender: para diferentes categorias de dados, para grande escala, para os com baixa densidade, valores e significados diversos. Podem também aprender para dados incertos e incompletos, e para alta velocidade de *streaming* (Singh & Jaiswal, 2018). Disponibiliza uma análise online, porque *big data* responde bem ao processamento em larga escala (Leskovec, Rajaraman, & Ullman, 2014). Normalmente, *machine learning* se utiliza de uma aprendizagem global, onde as suas estratégias são oriundas de uma análise generativa em que os dados resultantes do modelo podem gerar novas entradas. Basicamente, é um resumo de todo o conjunto de dados (K. Huang et al., 2005). Integrado aos sistemas de *big data*, *machine learning* tem a possibilidade de fazer uso de uma aprendizagem local, ou seja, disponibiliza uma estratégia que oferece uma análise apenas nos subconjuntos de interesse (Bottou & Vapnik, 1992). Portanto, é disponibilizar uma análise com hipóteses mais fortes e menos restritivas, o que gera maior rendimento ao modelo, baixa variância e polarização (Al-Jarrah, et al., 2015). Além disso, a aprendizagem local superou a global em precisão e tempo de computação em vários estudos de previsão (Elattar, Goulermas, & Wu, 2010; Grolinger, et al., 2016).

Do & Poulet, 2015 desenvolveu um algoritmo que executa em paralelo um conjunto de algoritmos de SVM locais e aleatórios. Isso permitiu nas abordagens relacionadas com o volume, um melhor desempenho em relação ao algoritmo SVM típico. Este modelo se utiliza do conceito de aprendizagem de transferência, ou seja, melhora a análise num domínio particular, denominado de domínio alvo, treinando o modelo com outros conjuntos de dados de vários domínios (domínios de origem) com atributos ou características semelhantes (Do & Poulet, 2015). É utilizada quando o tamanho dos dados dentro do domínio de destino for insuficiente ou de análises de diferentes tarefas (Torrey & Shavlik, 2010). Um exemplo de caso de utilização em crédito é a previsão de incumprimento de uma empresa nova (domínio de destino) utilizando conjuntos de dados recolhidos a partir de outras empresas semelhantes (domínios de origem) que provavelmente têm padrões similares, mas são diferentes em tamanho e eficiência.

Outra vantagem que *machine learning* pode proporcionar a gestão de risco de crédito bancário é uma verificação ao longo da vida, que imita a aprendizagem humana; é a análise contínua; O conhecimento é mantido e usado para resolver problemas diferentes. É a análise dirigida para maximizar a aprendizagem, em geral, para ser capaz de resolver uma tarefa, treinando tanto para um único domínio como para domínios heterogêneos (Silver, Yang, & Li, 2013). Como na análise *online*, a aprendizagem ao longo da vida é um processo contínuo. No entanto, enquanto a aprendizagem *online* considera apenas um único domínio, a aprendizagem ao longo da vida inclui uma multiplicidade de domínios. Já com relação à aprendizagem de transferência, a análise ao longo da vida se assemelha por ser capaz de transferir conhecimentos entre os domínios. Se difere por ser um processo de refinamento contínuo do tópico. Ou seja, a cada chegada de um novo resultado, ao longo do tempo, o tópico é refinado. Por exemplo, considere os processos de capacidade de cumprimento de um cliente quando estão disponíveis vários dados, como a sua localização física, as interações de redes sociais, dados meteorológicos e histórico de operações. A previsão de interesses individuais, precisa modificar periodicamente ao longo do tempo quando estiverem disponíveis novos dados, pois, a capacidade financeira do cliente pode mudar com o tempo.

Para finalizar esta secção, algoritmos de *machine learning* tradicionais têm um único domínio de destino e não podem transferir o aprendizado de uma tarefa para outra (com a exceção da análise de transferência). (Khan et al., 2016). Consequentemente, os autores consideram tais abordagens inadequadas para *big data* por causa dos muitos domínios abrangidos por esses dados e o aparecimento constante de novas informações. Eles consideram a aprendizagem ao longo da vida mais adequadas para soluções que se integram com sistemas de *big data* porque o modelo é continuamente refinado e a tarefa aprendida é aplicável a diferentes domínios.

2.3.2. Estudos Comparativos

Nesta seção apresentou-se na Tabela 7, a variedade de resultados de alguns estudos comparativos dos principais modelos de *machine learning* com foco para as situações semelhantes de risco de crédito bancário. Os valores que lá estão apresentados são relacionados a acurácia dos testes. Como acurácia entendeu-se como o grau de proximidade que uma estimativa tem de seu parâmetro, ou seja, proximidade do valor verdadeiro. (Mikhail e Ackermann citado por Monico et al., 2009).

A legenda utilizada foi:

Melhor resultado	Segundo melhor	Terceiro melhor	Quarto melhor	Quinto melhor	Apenas 1 resultado
------------------	----------------	-----------------	---------------	---------------	--------------------

Tabela 7 – Comparativo dos principais modelos de ML para crédito - 1996 a 2019

<i>Autores</i>	<i>RL</i>	<i>RNA</i>	<i>GA</i>	<i>DT</i>	<i>DA</i>	<i>SVM</i>
(Mendes Filho, Carvalho, & Matias, 1996)	-	96,6	-	-	-	-
(Arminger, Enache, & Bonne, 1997)	67,6	65,2	-	66,4	-	-
(Maher & Sen, 1998)	61,7	66,7	-	-	-	-
(Arraes, Semolini, & Picinini, 1999)	84,8	85,4	-	-	-	-
(Fritz & Hosemann, 2000)	-	81,6	82,4	79,5	82,7	-
(Rosa, 2000)	70,4	-	-	66,6	-	-
(Nanda & Pendharkar, 2001)	-	-	65,0	-	62,5	-
(M.C. Chen, Huang, & Chen, 2002)	-	91,9	92,9	-	-	-
(Semolini, 2002)	68,3	67,4	-	-	-	-
(I. A. Guimarães & Chaves Neto, 2002)	99,0	-	-	-	92,1	-
(Picinini, Oliveira, & Monteiro, 2003)	63,5	64,4	67,5	-	-	-
(J. -Z. Huang & Kong, 2003)	73,9	77,6	-	-	-	78,9
(Eifert, 2003)	88,2	-	-	-	-	-
(Z. Huang et al., 2004) *	74,0	76,9	-	-	-	78,9
(Li et al., 2004)	-	-	-	-	-	78,7
(Lemos, Steiner, & Nievola, 2005)	-	90,0	-	71,9	-	-
(Gonçalves, 2005)	68,3	67,7	66,5	-	-	-
(Y. Wang, S. Wang, & Lai, 2005) *	72,9	72,3	-	-	-	73,3
(Lai et al., 2006)	75,8	80,7	-	-	-	79,9
(Protil & Silva, 2006)	91,2	-	-	-	88,2	-
(Abu-Nimeh et al., 2007)	95,1	94,1	-	92,3	-	92,0
(Sabzevari, Soleymani, & Noorbakhsh, 2007)	72,5	66,0	-	82,5	-	-
(Selau, 2008)	84,1	83,5	-	-	83,9	-
(Pires, 2008)	-	96,4	-	-	-	-

Tabela 7 – Continuação.

(Ince & Aktan, 2009)	62,3	61,5		65,5	62,2	-
(W. Chen, Ma, & Ma, 2009)	-	-	-	81,7	-	81,2
(L. Yu, Wang, & Lai, 2009)	83,0	81,0	-	-	-	82,2
(Tsai & Chen, 2010) *	75,5	75,1	-	69,2	-	-
(H. Yu et al., 2010) *	79,0	75,0	-	77,0	-	78,6
(L. Yu et al., 2010)	64,6	58,8	-	-	60,8	65,2
(G. Wang et al., 2011)	78,2	75,2	-	78,0	-	76,4
(Bhattacharyya et al., 2011)	94,7	-	-	-	-	93,8
(Ghodselahe & Amirmadhi, 2011)	71,0	68,5	-	70,3	65,9	62,6
(Akkoç, 2012)	57,8	58,6	-	-	-	-
(Olson, Delen, & Meng, 2012)	81,3	79,8	-	94,8	-	-
(Swiderski, Kurek, & Osowski, 2012)	82,0	-	-	81,7	-	-
(Kambal et al., 2013) *	-	66,7	-	63,0	-	-
(Cubiles-De-La-Veja et al., 2013)	86,2	-	-	85,9	86,4	86,3
(Erdal & Ekinci, 2013) *	-	77,5	-	-	-	79,5
(Bekhet & Eletter, 2014)	85,4	81,5	-	-	-	-
(Mishra & Dash , 2014)	-	81,8	-	84,2	-	-
(Sousa & Figueiredo, 2014)	-	95,5	-	97,0	-	-
(Harris T. , 2015)*	70,4	76,1	-	-	-	75,2
(Koutanaei, Sajedi, & Khanabaei, 2015)	-	87,2	-	85,9	-	85,9
(Salgado, 2015)	81,5	83,2	-	-	-	-
(Tarihi & Tarihi, 2016)	-	-	-	-	-	83,8
(Turkson, Baagyere, & Wenya, 2016)	78,0	80,0	-	73,0	81,0	78,0
(Ala'raj & Abbod, 2016)*	-	77,9	-	82,3	-	79,3
(Bequé & Lessmann, 2017)*	77,0	74,7	-	60,1	-	75,1
(Barboza, Kimura, & Altman, 2017)	82,7	84,8	-	-	64,8	88,75
(Luo, Wu, & Wu, 2017)	77,3	87,7	-	-	-	87,4
(H. He, Zhang, & Zhang, 2018)	78,7	68,8	-	64,7	78,8	75,2
(Rajamohamed & Manokaran, 2018)	-	76	-	83,7	-	93,2

Tabela 7 – Continuação.

(Sabbeh, 2018)*	86,7	89,5	-	87,9	85,3	92,6
(Arora, Fan, & Leblanc, 2019)	59	-	-	61	-	-
(Munkhdalai et al., 2019)*	79,2	-	-	-	-	70,9
(Amanze, Asogwa, & Chukwunke, 2019)	-	81,8	-	82,0	-	-

* Média de todos os *datasets* existentes na citação.

Fonte: Elaboração própria.

Ao analisar a Tabela 7, por se tratar de fonte de dados diferentes, não se pôde comparar uma citação com a outra. Porém, consegue-se perceber quais os modelos que estão com maior frequência nos estudos da academia e quais deles estão com melhores resultados individuais. Antes de analisar profundamente este resultado, aplicou-se a grelha de pontuação da Tabela 8, para uma melhor evidência.

Tabela 8 – Grelha de pontuação para avaliação dos modelos de ML para crédito

<i>Melhor resultado</i>	<i>Segundo melhor</i>	<i>Terceiro melhor</i>	<i>Quarto melhor</i>	<i>Quinto melhor</i>
5 pontos	4 pontos	3 pontos	2 pontos	1 ponto

Fonte: Elaboração própria.

A Figura 3 detalha os resultados desses modelos que tiveram as melhores classificações.

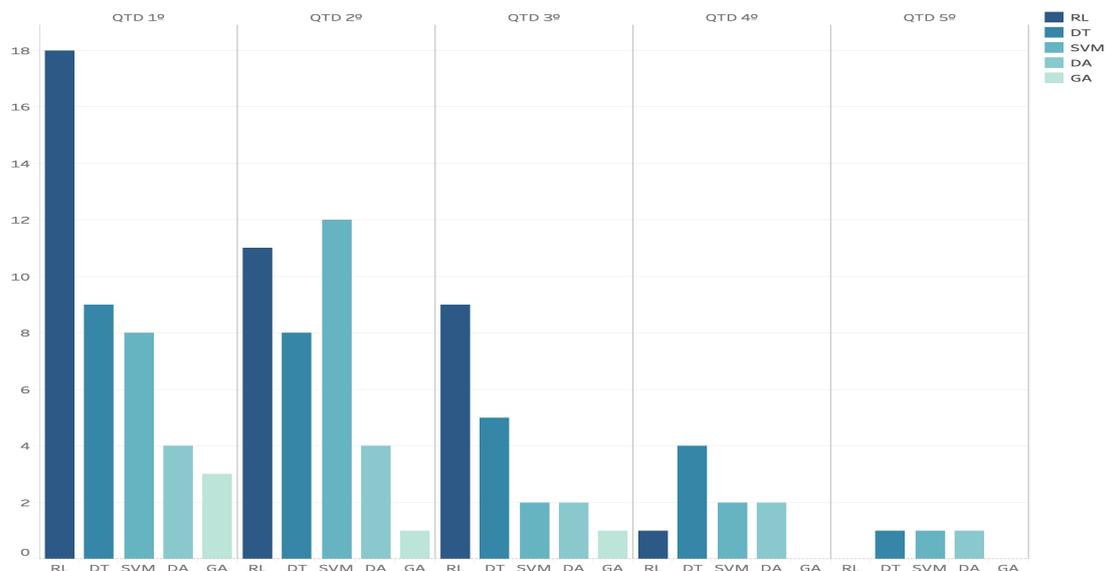


Figura 3 – Detalhamento dos resultados dos modelos da amostra.

Fonte: Elaboração própria.

Na análise da Tabela 7 já se conseguia prever que o modelo que possui um maior destaque é o de RL. É o modelo que possui o melhor resultado em mais de 15 citações,

em 11 citações foi o segundo modelo com melhor acurácia e em quase 10 citações foi o terceiro melhor modelo. Com a aplicação da grelha de pontuação da Tabela 8, formou a seguinte classificação:

Tabela 9 – Classificação da amostra de modelos de ML para crédito

Modelos de ML para Risco de Crédito	Pontos	Classificação
RL – Regressão logística	163	1º
RNA – Rede neural artificial	159	2º
DT – <i>Decision trees</i>	100	3º
SVM – <i>Support Vectors Machine</i>	99	4º
DA – <i>Discriminant analyses</i>	47	5º
GA – <i>Genetic algorithm</i>	22	6º

Fonte: *Elaboração própria.*

Por fim, mostrou-se na Figura 4, um estudo na literatura, onde o autor demonstrou a preferência de modelos de ML em risco de crédito devido, principalmente, a sua capacidade preditiva. Onde uma vez mais a regressão logística destacou-se em 1º lugar:



Figura 4 – Percentual de preferência de modelos de ML.

Fonte: (R. A. Silva, 2014).

Capítulo 3 – Metodologia

Para analisar o uso de *machine learning* nos processos de gestão de risco de crédito das instituições bancárias, realizou-se inicialmente uma pesquisa exploratória com intuito de adquirir mais familiaridade sobre o tema. Foram investigadas referências na bibliografia em três nichos: O primeiro sobre o termo “risco de crédito”; O segundo sobre “*machine learning*” e o terceiro nicho foi a fusão dos dois anteriores: “*machine learning* para risco de crédito”. Os principais autores, que contribuíram com o trabalho, foram Dimitri Bertsekas, Jared Dean, John McCarthy, Kevin Knight, Peter Norvig e Stuart Russell.

Após esta fase inicial, foram feitas investigações descritivas, através de dois questionários, aos profissionais de crédito das instituições bancárias elencadas para este estudo. O primeiro, mais amplo, tiveram questões mais técnicas. O segundo foi feito um refinamento dos resultados obtidos do primeiro questionário e disponibilizados aos executivos e diretores de créditos com intuito de validação.

Quanto às instituições bancárias foram escolhidas dentre três grupos específicos: os principais bancos do mundo, maiores do Brasil e alguns dos maiores de Portugal. Por razão de confidencialidade, os nomes dos bancos tiveram que ser omitidos. Sendo assim, a amostra consiste em 12 instituições bancárias: três do Brasil, duas de Portugal, duas dos EUA, duas da China, uma de Espanha, uma do Reino Unido e uma de França.

Dado à impossibilidade de contato perante a sede, dado que numa primeira abordagem não houve resposta, no primeiro questionário foram abrangidos os diversos funcionários destes bancos selecionados na internet, através da rede, LinkedIn.

Este primeiro questionário, tiveram 26 questões. Divididas na sua maioria por questões de múltipla escolha, questões de classificação (ordenação) e questões de escala com mensuração, tipo Likert de 5 opções. As questões de múltiplas escolhas tiveram algumas do tipo, fechada, onde os candidatos só podem escolher uma das alternativas pré-definidas. Outras semiabertas, onde o entrevistado pudera escolher mais de uma opção, além de incluir outras opções de escolha. As questões de classificação também foram de tipo fechada e semiabertas como as de múltipla escolha. As respostas às questões do tipo Likert são pontuadas de -2 a 2: 2 “concordo totalmente”; 1 “concordo”, 0 “indiferente”, -1 “discordo”, -2 “discordo totalmente”.

Esse instrumento foi previamente desenvolvido e transportado para uma plataforma online (*Survio*) visando facilitar o acesso e participação. O questionário foi de resposta anónima e participação voluntária, tendo estado disponível durante os meses de janeiro a março de 2019.

Foi considerado na análise dos resultados, quatro categorias de questões, destinadas a avaliar diferentes aspetos:

- Perfil do entrevistado (afirmações 1 a 4, 8);
- Questões relacionadas à empresa (afirmações 5 a 7);
- Questões de *machine learning* (afirmações 9 a 18);
- Questões de *machine learning* para risco de crédito (afirmações 19 a 26).

No segundo questionário, como o intuito era de validação, foram refinados os resultados do primeiro em 11 questões. Sendo que quase na sua totalidade, foram questões cujos resultados eram “sim” ou “não”.

A abordagem adotada foi do tipo quantitativo, onde para se chegar aos resultados foram usados os números extraídos dos questionários *online*.

O estudo também foi reforçado com uma análise bibliométrica. Nesta análise foram utilizados dados dos principais sites de pesquisa da literatura académica: b-on, web of science, scopus e google académico. Aqui como critério de mineração foram utilizados o terceiro nicho (*machine learning* para risco de crédito”) e a seleção dos primeiros 100 resultados de cada site de pesquisa mencionado anteriormente.

A pesquisa foi realizada no Instituto Universitário de Lisboa numa amostra de 12 instituições bancárias. Dentre elas possuem instituições brasileiras, portuguesas e alguns dos principais bancos do mundo.

Quanto aos procedimentos metodológicos, este estudo caracterizou-se como documental. Onde se buscou informações através da análise de documentos de sites, softwares, revistas, jornais e livros.

Com relação à análise dos dados de uso de *machine learning* nos processos de gestão de risco de crédito das instituições bancárias, nesse ponto, caracterizou-se como quantitativa.

Capítulo 4 – Análise e Discussão dos Resultados

Neste capítulo, visa-se partilhar os resultados do estudo empírico composto pela recolha e análise das respostas dadas ao questionário virtual dirigido às instituições bancárias.

4.1. Análise Bibliométrica

Iniciou-se a análise bibliométrica pesquisando no Google Académico, o perfil dos autores que foram citados neste estudo cuja afinidades de suas obras estão mais relacionadas a *machine learning* e crédito. A Figura 5 demonstra o total de citações por autor.

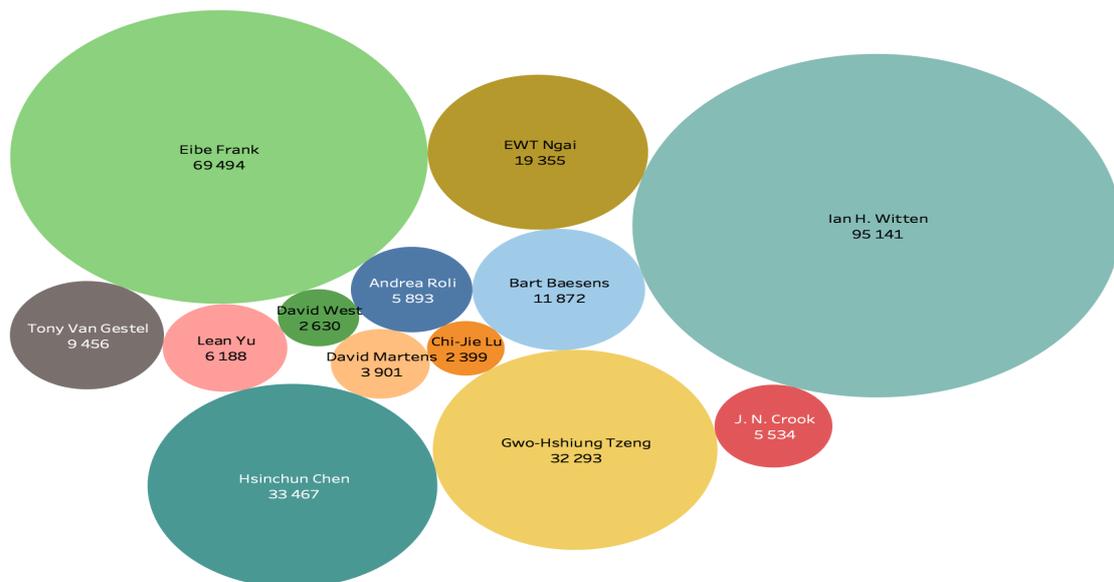


Figura 5 – Principais autores de machine learning para crédito

Fonte: Elaboração própria.

Ao analisar a Figura 5 percebe-se um destaque maior de Ian H. Witten com 95141 citações e Eibe Frank com 69494. A obra de ambos com mais citações foi: “*Data Mining: Practical machine learning tools and techniques*” com mais de 30.000 citações. Neste livro encontra-se conceitos, técnicas, métodos, ferramentas de *machine learning*, bem como sua aplicabilidade. O autor reporta que ML fornece a base técnica da DM. Ele é usado para extrair informações dos dados brutos em bancos de dados. Ou seja, informações, idealmente, expressas de forma compreensível que podem ser usadas para uma variedade de propósitos..., mas não há magia em ML, nenhum poder oculto, nenhuma alquimia. Em vez disso, há um corpo identificável de técnicas simples e práticas que muitas vezes podem extrair informações úteis de dados brutos” (Witten et al., 2017).

Sendo o terceiro mais citado, o Hsinchun Chen com 33467 citações. Neste destacamos a seguinte obra: “*Credit rating analysis with support vector machines and neural networks: a market comparative study*”. Onde o autor faz um estudo de comparação entre os modelos SVM e RNA. Mas antes menciona que: “Existem dois tipos básicos de notações de crédito, um para emissões de dívida específicas ou outras obrigações financeiras e o outro para emitentes de dívida...Ambos os tipos de notação são muito importantes para a comunidade de investidores. Uma notação mais baixa indica geralmente um risco mais elevado, o que tem um efeito imediato no rendimento de juros subsequente da emissão de dívida” (Z. Huang et al., 2004).

Outra afirmação importante, que se encontra na obra de Chen é: “Acredita-se geralmente que o processo de notação de crédito envolve uma avaliação altamente subjetiva dos fatores quantitativos e qualitativos de uma determinada empresa. As agências de rating e alguns pesquisadores enfatizaram a importância do julgamento subjetivo no processo de classificação de risco e criticaram o uso de modelos estatísticos simples e outros modelos derivados de técnicas de AI para prever as classificações de crédito.” Porém, a revisão da literatura do mesmo autor demonstrou que “os modelos estatísticos e os modelos de IA (principalmente redes neurais) alcançaram um desempenho de previsão notavelmente bom e captaram amplamente as características do processo de classificação de risco” (Z. Huang. et al., 2004).

Ainda sobre autores, a Tabela 10 elencou outros autores importantes selecionados nesta análise bibliométrica pelos seus conhecimentos aprofundados de *machine learning*, gestão de risco ou áreas correlacionadas a *machine learning* para crédito – *big data* e infraestrutura de tecnologia.

Tabela 10 – Outros autores importantes

Especialidade	Autores	Total de citações
Big Data	H. V. Jagadish	34.233
	J. Gehrke	39.343
Gestão de Risco	CH Wang	50.272
	G. A. Holton	1.433
Machine Learning	C. Pal	41.678
	M. Hall	74.009
	P. Norvig	48.667
	P. Tamayo	75.247
	S. Russell	71.489
Tecnologia da Informação	T. Davenport	93.270
	I. Polato	156
	Laat	5.696

Fonte: Elaboração própria.

Em continuidade na análise bibliométrica, mas com foco nos artigos, a Tabela 11 listou importantes artigos relacionados a *machine learning* para crédito.

Tabela 11 – Principais Artigos

Título do artigo	Autores	Total de citações	Citações desde 2015
<i>Credit rating analysis with support vector machines and neural networks: a market comparative study</i>	(Z. Huang et al., 2004)	943	641
<i>Neural network credit scoring models</i>	(West, 2000)	968	622
<i>Credit scoring with a data mining approach based on support vector machines</i>	(C. -L. Huang, Chen, & Wang, 2007)	666	381
<i>Credit scoring using the hybrid neural discriminant technique</i>	(T. -S. Lee et al., 2002)	382	271
<i>A New Fuzzy Support Vector Machine to Evaluate Credit Risk</i>	(Y. Wang, S. Wang, & Lai, 2005)	361	231
<i>Using neural network ensembles for bankruptcy prediction and credit scoring</i>	(Tsai & Wu, 2008)	377	229
<i>Comprehensible credit scoring models using rule extraction from support vector machines</i>	(Martens et al., 2007)	380	228
<i>Building credit scoring models using genetic programming</i>	(Ong, Huang, & Tzeng, 2005)	331	219
<i>Genetic algorithms applications in the analysis of insolvency risk</i>	(Varetto, 1998)	280	212
<i>Hybrid mining approach in the design of credit scoring models</i>	(Hsieh, 2005)	196	147
<i>Application of support vector machines to corporate credit rating prediction</i>	(Y.-C. Lee, 2007)	188	136
<i>An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring</i>	(L. Yu, Wang, & Lai, 2009)	199	134
<i>Feature selection in bankruptcy prediction</i>	(Tsai, 2009)	204	126

A Figura 8, confirma a importância dos três principais termos, e o principal é que estão intrinsecamente relacionados dentro da literatura.

4.2. Questionário – Aos Técnicos de Crédito

4.2.1. Recolha de Dados

Relativamente a recolha de dados do questionário “Aos Técnicos de Crédito”, a amostra consiste em 73 respostas dos 517 profissionais de créditos selecionados das instituições bancárias citadas no capítulo anterior. E este foi elaborado com o objetivo de avaliar o contexto do *machine learning* em risco de crédito de instituições bancárias.

4.2.2. Resultados

Iniciou-se esta secção de resultados com relação às informações do questionário. Aqui descobriu os países com mais participação, distribuição das respostas por bancos e a duração de preenchimento do questionário. Para este levantamento foram feitas as Figuras 9 a 11, mensurada por instituições bancárias. Por questões de anonimato e sigilo acordado com os entrevistados, os bancos foram identificados por numeração ordinal.

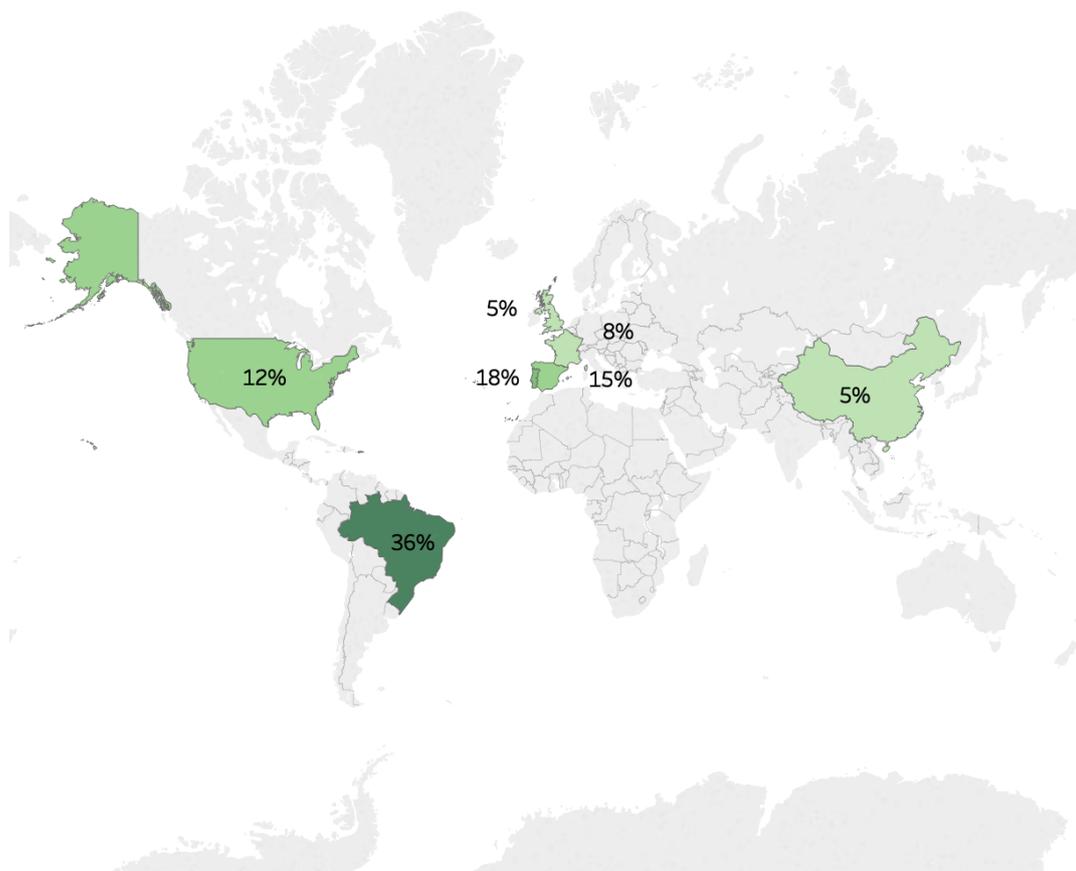


Figura 9 – Distribuição das respostas do questionário aos técnicos de crédito.

Fonte: Elaboração própria.

A Figura 9, demonstrou que houve um maior número de respostas vindas das instituições bancárias brasileiras. Seguidas das instituições portuguesas e espanholas. Como fator negativo, o baixo índice de respostas das instituições do Reino Unido e China.

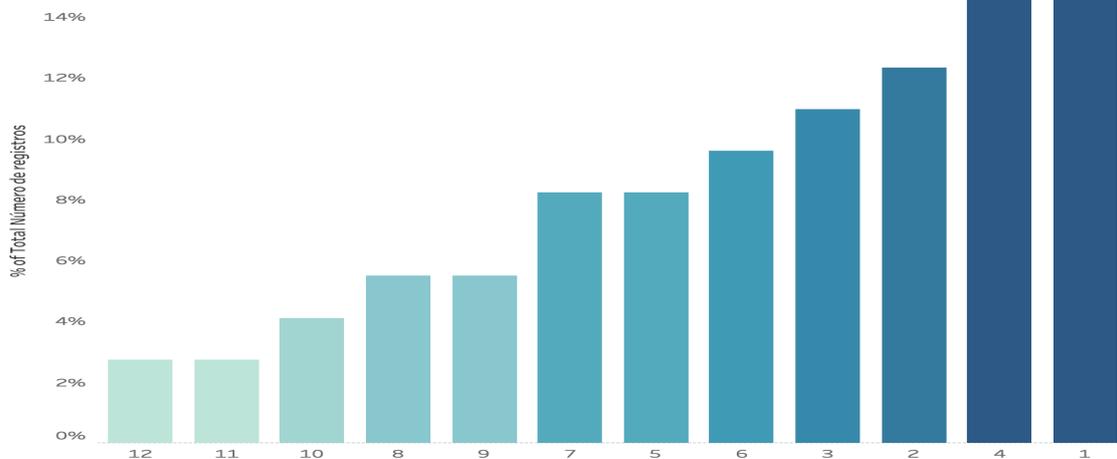


Figura 10 – Percentual de respostas por banco.

Fonte: Elaboração própria.

Já a Figura 10, procurou aprofundar ainda mais a análise das informações obtidas com o questionário “Aos técnicos de crédito” e identificar os bancos que mais participaram deste estudo. Percebeu-se um destaque para os bancos 1 e 4 com mais de 15% cada. Seguidos pelos bancos 2, 3 e 6. Estes juntos tem mais da metade das respostas.

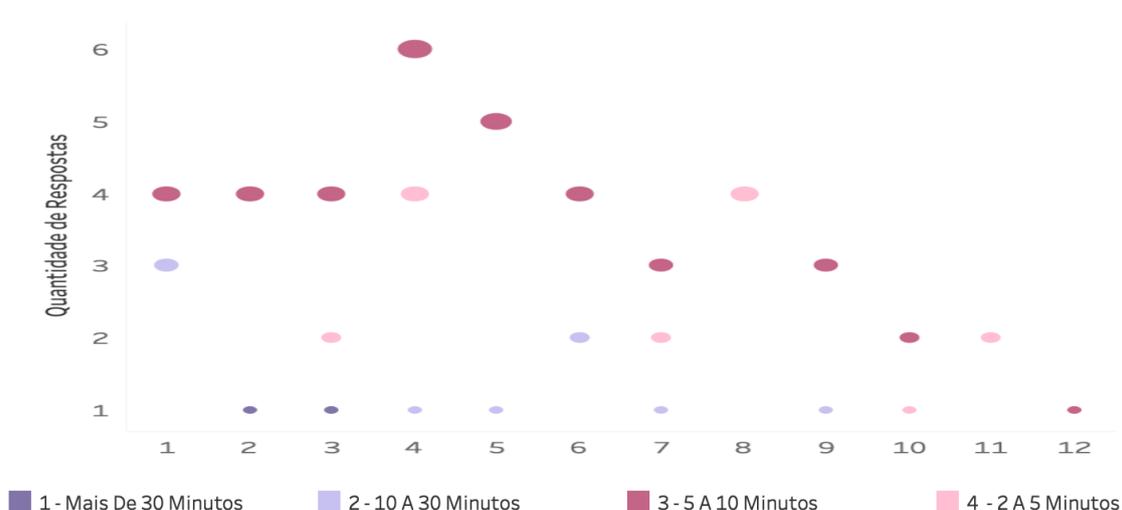


Figura 11 – Desempenho de respostas por banco.

Fonte: Elaboração própria.

Com análise da Figura 11, pode-se inferir que o questionário foi respondido de forma ágil, onde a grande maioria dos entrevistados (mais de 80%) levaram até 10 minutos para responderem ao questionário.

A seguir iniciou-se a análise do perfil da amostra de pessoas pesquisadas em relação aos cargos em que se enquadram os entrevistados:

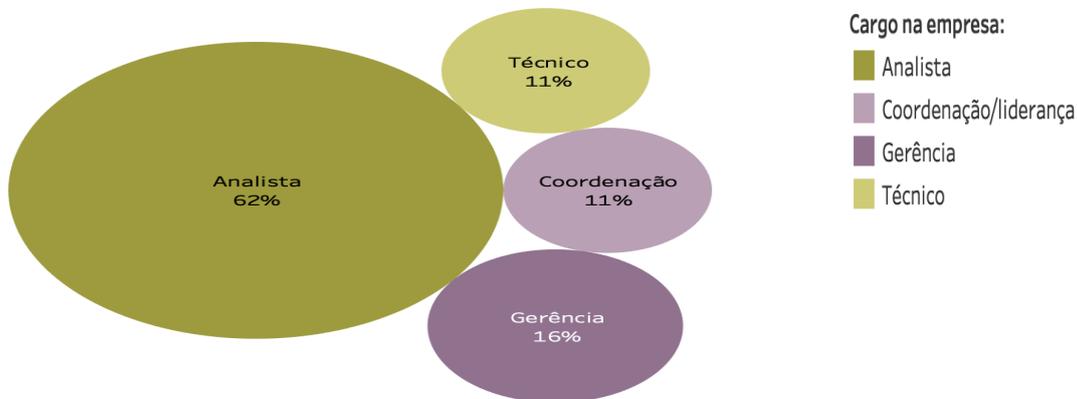


Figura 12 – Tipo de cargo que exerce o entrevistado.

Fonte: Elaboração própria.

Consoante a análise da Figura 12, não existe nenhum entrevistado do tipo “direção”. Já do tipo “gerência” existem 16% dos entrevistados, o tipo “coordenação ou liderança” alcançou 11 % dos entrevistados, o tipo “Analista” atingiu 62% e o tipo “técnico” ficou com 11%. Com isso, pode-se concluir que a maioria destes entrevistados possuem o cargo de analista. Em virtude disso e de um melhor entendimento, as análises a seguir mencionaremos o cargo como dois grandes blocos: técnico (analistas e técnicos) e gerência (gerentes, coordenadores e líderes).

Com a Figura 13 pode-se vislumbrar as informações da amostra em relação com a idade e cargo:



Figura 13 – Faixa de idade do entrevistado.

Fonte: Elaboração própria.

Segundo a Figura 13, a faixa etária “até 25 anos” possui 14% de respostas, a “de 25 a 35 anos” ficou com 41% de respostas, a “de 35 a 45 anos” alcançou 31% de respostas, e a “de 45 a 60 anos” atingiu 13%, podendo concluir-se que a maioria destes entrevistados possuem uma idade entre 25 a 45 anos. Na correlação com o cargo, percebe-se que na

amostra deste estudo, todas as faixas etárias têm um maior número de analista do que gerente. Exceto na faixa “de 45 a 60 anos” que os números são próximos.

Na Figura 14, obteve-se as informações de escolaridade dos entrevistados:

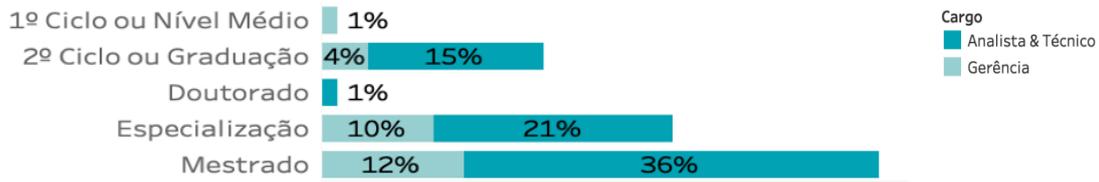


Figura 14 – Nível de graduação do entrevistado.

Fonte: *Elaboração própria.*

Dentre os entrevistados da análise da Figura 14, 1% dispõe da escolaridade de “doutorado”, 48% “mestrado”, 31% “especialização”, 19% “graduação” e 1% o nível “médio”. Com isso, pode-se concluir que a maioria destes entrevistados possuem um título de mestrado ou especialização. Na correlação com o cargo, a amostra deste estudo traz uma informação que não é comum com a realidade: O nível maior de escolaridade ser analista e o menor ser gerente. Como são percentuais baixo (2 % da amostra) não foi considerado um dado importante que pudesse afetar o resultado final. Considerou a informação principal que na maioria das faixas de escolaridade tiveram o cargo de analista como predominante.

Na Figura 15 pode-se analisar o tempo de empresa que os entrevistados têm e sua relação com o cargo:

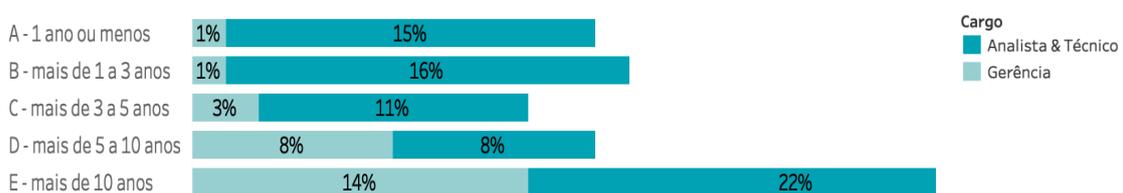


Figura 15 – Tempo de empresa do entrevistado.

Fonte: *Elaboração própria.*

De acordo com a Figura 15, o entrevistado com “1 ano ou menos” de empresa foi de 16%, o com “1 a 3 anos” de empresa foi 17 %, o “3 a 5 anos” de empresa foi 14%, o com “5 a 10 anos” de empresa foi 16% e o com “mais de 10 anos foi de 36%. Com isso, pode-se concluir que a maioria destes entrevistados estão a mais de 10 anos em suas empresas. A correlação aqui das faixas que representam mais tempo de empresa apresentam um maior equilíbrio entre os técnicos e gerentes.

Na Figura 16, pode-se analisar se o entrevistado trabalha com alguma tecnologia relacionada ao tema deste estudo (*machine learning*, *big data*, ou infraestrutura de tecnologia):

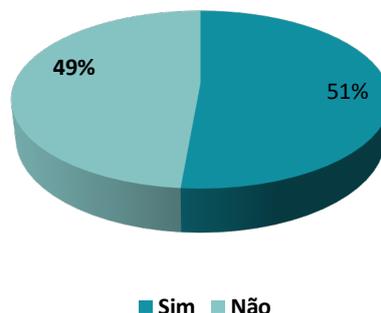


Figura 16 – Você participa de alguma atividade relacionada a ML, Big Data ou TI?

Fonte: *Elaboração própria.*

O “sim” obteve 51% das respostas e o “não” obteve 49% segundo exposto na Figura 16. Com isso, pode-se concluir que a amostra é bem dividida com uma leve maioria destes entrevistados estarem a trabalhar com uma das tecnologias deste estudo.

Com as afirmações relacionadas à empresa tentou observar as características das empresas dos entrevistados, para que se evite comparar empresas de portes diferentes. Na avaliação desta análise teve-se que considerar que alguns dos entrevistados por serem novos nas empresas, podem não saber com exatidão a dimensão de suas empresas. Portanto, um critério utilizado foi que quando uma empresa tiver duas respostas distintas para as afirmações relacionadas às características, leva-se em conta os dados da maioria das respostas daquela empresa e depois daqueles que têm mais tempo de empresa.

Iniciou-se esta análise com a quantidade de funcionários que as empresas dos entrevistados têm, conforme Figura 17:



Figura 17 – Quantidade de funcionários das empresas dos entrevistados.

Fonte: *Elaboração própria.*

Dentre os entrevistados da Figura 17, 88% responderam que suas empresas possuem “mais de 1000” funcionários. E 5% dos entrevistados responderam que suas empresas possuem “mais de 100 até 500” funcionários e os 7% restantes responderam que suas

empresas possuem “mais de 500 até 1000” funcionários. Com isso, pode-se concluir que a maioria das empresas são de porte de mais de 1000 funcionários. Este resultado é positivo porque a amostra possui a quantidade desejada de funcionários para o estudo.

Na Figura 18 pode-se analisar a empresa de acordo com seu volume de negócios. O volume de “mais de 50” milhões de euros foi apresentado em 100% das respostas da grande maioria das instituições bancárias. Apenas um banco português e outro do nicho dos internacionais que tiveram 89% e 91% respectivamente.



Figura 18 – Volume de negócio das empresas dos entrevistados.

Fonte: Elaboração própria.

Conclui-se com isso, que a maioria dos bancos ilustrados na Figura 18, são de porte negocial de mais de 50 milhões de euros. O que torna a amostra factível para o estudo.

Na análise da empresa de acordo com seu volume de negócios em crédito, em razão do valor desejável para este estudo serem acima de 50% e uma melhor visualização, aglutinou-se os resultados em apenas dois níveis:



Figura 19 – Volume de negócio em crédito das empresas dos entrevistados.

Fonte: Elaboração própria.

De acordo com os entrevistados ilustrados na Figura 19, a maioria das empresas são de porte negocial em crédito de mais de 50% de seus negócios e apenas 15 % responderam que suas empresas possuem “até 50%”. Resultado este que não invalidou a amostra em questão.

Após estas explicações a respeito das características das instituições entrevistadas, passou-se para a análise das percepções dos entrevistados. A partir deste momento, com intuito de enriquecer a análise, decidiu-se selecionar as respostas das instituições por países, agrupando-os em três grupos: Brasil, Portugal e Internacionais. Além disso, foi agrupado as respostas em três nichos: Concorda, Discorda e Indiferente. A primeira foi com relação aos benefícios de *machine learning*, conforme a Figura 20:

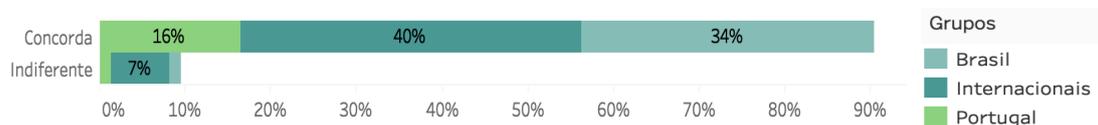


Figura 20 – *Machine learning* trará benefícios aos negócios de sua empresa?

Fonte: Elaboração própria.

Consoante com as respostas dos entrevistados da Figura 20, nenhuma pessoa discordou desta questão. Apenas 9% de percepções indiferentes. Já nas respostas positivas, 90% de percepções positivas e a resposta média foi de 4,3. Com isso, pode-se concluir que a maioria das empresas acreditam que *machine learning* trará benefícios significantes aos negócios da empresa.

Na Figura 21, pode-se analisar a presença de *machine learning* estará entre os novos produtos:

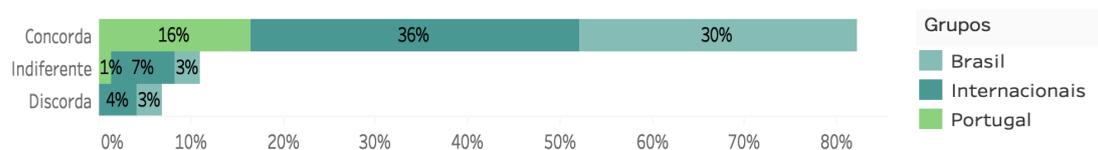


Figura 21 – *Machine learning* estará presente em todos os novos produtos e serviços?

Fonte: Elaboração própria.

Na análise da Figura 21, 11% discordaram desta questão, 7% foram “indiferentes”, gerando um percentual parcial de 18% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 82% de percepções positivas e a resposta média foi de 4. Outro ponto a mencionar é que o único grupo que não teve discordância neste assunto foi o das instituições portuguesas. Diante disso, pode-se concluir que a maioria das empresas acreditam que *machine learning* estará presente em quase todos os novos produtos e serviços de softwares adquiridos.

Em seguida, conforme as Figuras 22 e 23, analisa-se o quanto das instituições bancárias estão a usar *machine learning*:

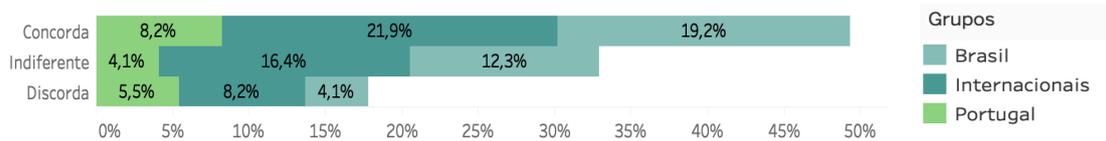


Figura 22 – Considera que sua empresa atua em machine learning?

Fonte: Elaboração própria.

Nesta questão da Figura 22, 17,8% discordaram, 32,8% foram indiferentes, gerando um percentual de 50,6% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 49,4% percepções positivas e a resposta média foi de 3,4. Assim, é possível concluir, que de acordo com os dados, a metade das instituições bancárias já atuam com *machine learning*. Além disso, é bom mencionar o alto percentual de respostas como indiferentes. O que pode ser um indicativo que parte dos entrevistados não tinha muita certeza sobre o que responder nesta questão.

Após isso, passou-se para a análise da intenção de implementação de *machine learning* nas instituições bancárias até 2020, conforme ilustração da Figura 23:

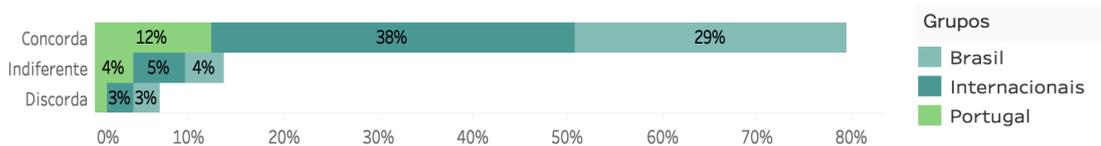


Figura 23 – Pensa em investir em pelo menos um projeto de ML até 2020?

Fonte: Elaboração própria.

Na análise da Figura 23, apenas 7% discordaram e 13% dos entrevistados foram indiferentes, gerando um percentual parcial de 20% de percepção indiferente ou negativa a questão. Já nas respostas positivas, aproximadamente 80% de percepções positivas e a resposta média foi de 4,2. Portanto, é possível afirmar que a maioria das instituições bancárias pensam em implementar até 2020, pelo menos um projeto com ML.

A seguir na Figura 24, foi feita a análise da percepção dos bancos sobre as capacitações em ML. Se acham que é um problema ou não para encontrarem profissionais:

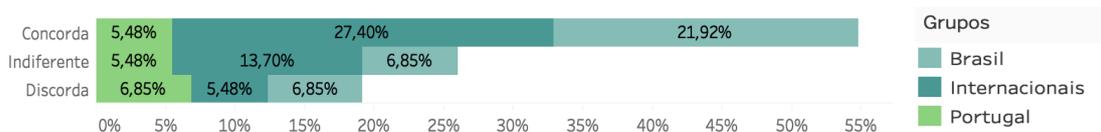


Figura 24 – Falta de profissionais em ML é um dos problemas da sua empresa?

Fonte: Elaboração própria.

Aqui na Figura 24, obteve-se 19% dos entrevistados que discordaram desta questão, 26% foram indiferentes, gerando um percentual parcial de 45% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 55 % dos entrevistados concordaram e a resposta média foi de 3,6. Então, pode-se concluir que um pouco mais da metade dos bancos acham ser um problema a escassez de profissionais qualificados em *machine learning*. Aqui vale ressaltar também é que o número de indiferentes foi um tanto alto, o que prejudicou a pontuação individual da questão.

Na Figura 25, analisa-se a percepção sobre os executivos das instituições bancárias com relação aos benefícios de *machine learning*:

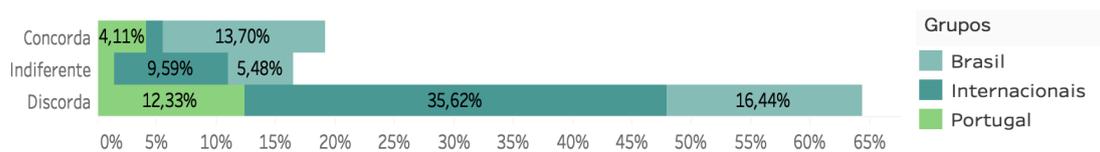


Figura 25 – A maioria dos executivos compreende os benefícios de ML?

Fonte: Elaboração própria.

Pela primeira vez neste questionário, uma questão onde o grupo de discordantes é maior que os que concordam. A Figura 25 apresentou que 64% dos entrevistados discordaram desta questão, 16% foram indiferentes, gerando um percentual de 80% de percepção indiferente ou negativa a questão. Já nas respostas positivas, aproximadamente 20% percepções de concordância e a resposta média foi de 3,7. Portanto, pode-se concluir que a maioria das instituições bancárias pensam que seus executivos não compreendem os benefícios do *machine learning*.

Em compreendendo os benefícios do *machine learning*, passa-se a analisar se os executivos dos bancos acreditam que esta mudança de mentalidade é um processo longo e lento, conforme a ilustração da Figura 26:

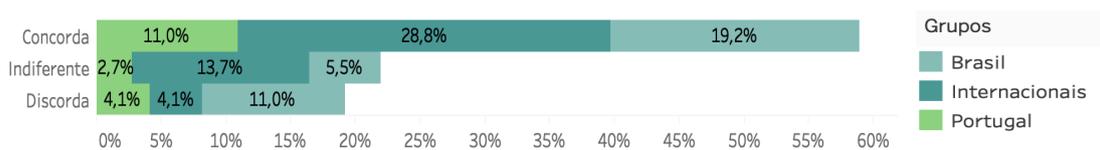


Figura 26 – Os executivos acreditam que ML é um processo longo e lento?

Fonte: Elaboração própria.

Dentre a amostra da Figura 26, cerca de 19% das pessoas discordaram desta questão, 21% das pessoas foram indiferentes, gerando um percentual de 40% de percepção

indiferente ou negativa a questão. Para a concordância da questão o percentual aproximado foi de 60% e a resposta média foi de 3,5. Dessa forma, apesar do alto percentual de respostas como indiferentes, infere-se que a maioria dos executivos das instituições bancárias acreditam que a mudança para *machine learning* é um processo longo e lento.

Na Figura 27, passou-se para uma análise com relação aos dados. Se as instituições bancárias acham que antes de implementar um projeto de *machine learning*, os processos com dados devem ser melhorados:

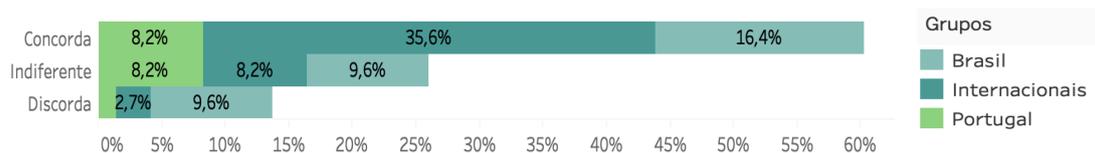


Figura 27 – Precisam melhorar os processos a nível de dados para depois usar ML?

Fonte: Elaboração própria.

Nesta questão da Figura 27, 13,7% dos entrevistados discordaram, 26% foram indiferentes, gerando um percentual de 39,7% de percepção indiferente ou negativa a questão. Já nas respostas positivas, foram 60,3% de concordância e a resposta média foi de 3,7. Em virtude disso, pode-se afirmar que a maioria das instituições bancárias acham que seus processos de dados devem ser melhorados antes de partirem para uma utilização plena de *machine learning*.

Em seguida, analisou-se as instituições bancárias com o foco na prestação dos serviços de *machine learning*. Se os bancos têm como estratégia terceirizar os serviços de TI de *machine learning*, conforme ilustração da Figura 28:

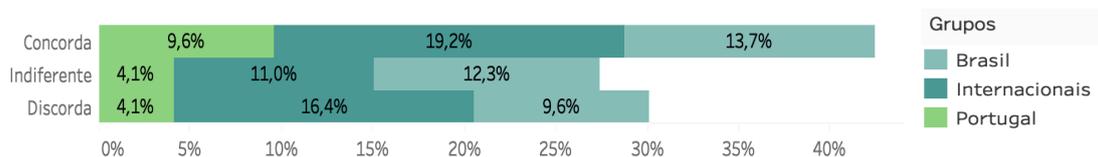


Figura 28 – Terceirização de TI para ML é uma das estratégias da sua empresa?

Fonte: Elaboração própria.

De acordo com as respostas dos entrevistados ilustradas na Figura 28, 30,1% dos entrevistados responderam em discordância a questão e 27,4 % dos entrevistados foram indiferentes, gerando um percentual parcial de 57,5% de percepção indiferente ou negativa a questão. Já nas respostas positivas, apenas 42,5% de percepções de concordância a

questão e a resposta média foi de 3. Com isso, pode-se concluir que a maioria das instituições bancárias não pensam que a terceirização dos serviços de *machine learning* como estratégia. Importante salientar também é que o maior número de entrevistados foi para a resposta “concordo parcialmente”. O que nos leva a pensar que apesar da maioria dos bancos não acharem estratégico, alguns já estão a pensar nesta possibilidade.

Por fim, foi feita uma questão de classificação, para analisar quanto a preferência das instituições bancárias em relação a que área pretendem desenvolver seus projetos de *machine learning*, conforme ilustrado na Figura 29:

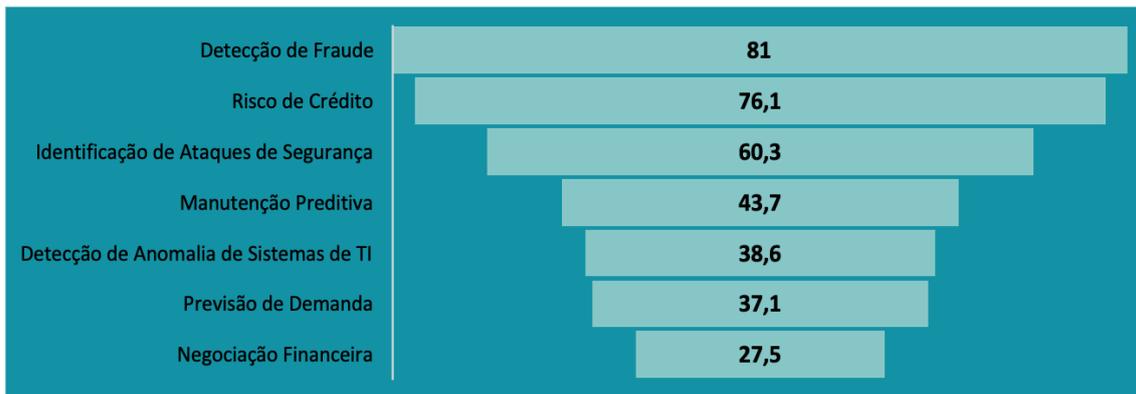


Figura 29 – Qual a área preferida para desenvolvimento em projetos de ML?

Fonte: Elaboração própria.

Nesta análise da Figura 29, percebe-se que a maioria das instituições financeiras têm uma preferência em desenvolver projetos de ML para “detecção de fraude” – 81 pts, “risco de crédito” – 76,1 pts e “identificação de ataques de segurança” – 60,3 pts.

A próxima seção avançou a análise para as afirmações de *machine learning* com foco no risco de crédito.

Neste ponto da análise teve-se como objetivo observar as instituições bancárias com foco em *machine learning* especificamente voltado para a área de risco de crédito.

Iniciou-se esta análise com a avaliação se a área de risco de crédito dos bancos dos entrevistados já possui algum projeto em *machine learning*, conforme Figura 30:

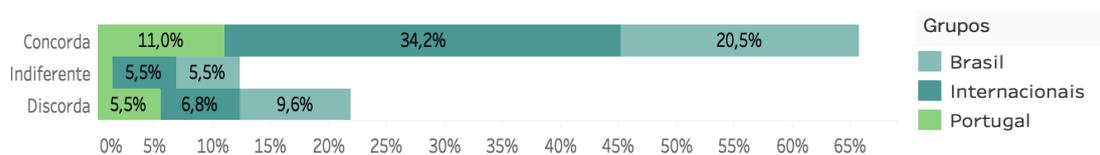


Figura 30 – O setor de risco de crédito já se utiliza de ML ou pensa em usar?

Fonte: Elaboração própria.

Entre os entrevistados, conforme a Figura 30, 21,9% responderam que discordam da questão, e 12,3% foram indiferentes, gerando um percentual parcial de 34,2% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 65,8% dos entrevistados responderam que concordaram e a resposta média foi de 3,7. Dessa maneira, a maioria das instituições bancárias consideram que seu setor de risco de crédito já se utiliza ou possui um potencial forte para utilizar *machine learning*.

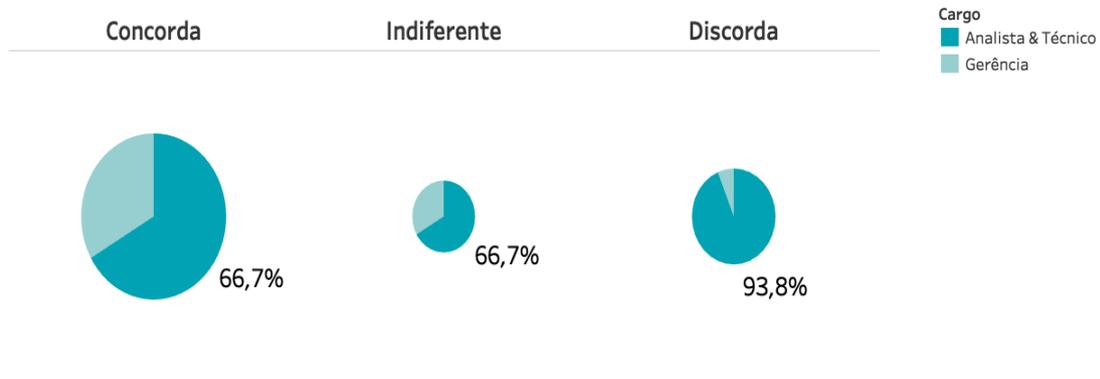


Figura 31 – Análise cruzada da atuação em Machine Learning por cargo.

Fonte: Elaboração própria.

Ao aprofundar na análise da atuação de ML no RC, percebeu-se através da Figura 31, que daqueles que concordaram com a questão, 66,7% são técnicos e que apenas 6,2% dos que discordaram são gerentes.



Figura 32 – Análise aprofundada do uso de Machine Learning por cargo.

Fonte: Elaboração própria.

Ademais, a Figura 32 esboça que tanto os gerentes quanto os técnicos, em sua maioria concordam que o setor de risco de crédito de suas empresas já se utiliza de *machine learning* ou pensa em usar. Cabe salientar que os gerentes possuem um percentual mais alargado – 80%. Enquanto no nível técnico a diferença percentual é menor, com 60,4% em concordância. Portanto, conclui-se que as instituições bancárias ainda podem disseminar melhor o conhecimento de *machine learning* junto ao seu nível técnico.

Em seguida, analisou se os analistas de crédito das instituições bancárias podem ser auxiliados por *machine learning* em suas tarefas diárias de análises e concessões de crédito, conforme ilustração da Figura 33:

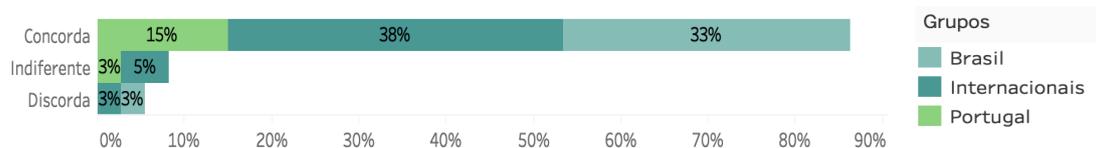


Figura 33 – ML pode auxiliar os analistas nas análises e concessões de crédito?

Fonte: Elaboração própria.

Nesta análise da Figura 33, nenhuma instituição portuguesa da amostra discordou desta questão, 6% responderam em discordância e 8% dos entrevistados foram “indiferentes”, gerando um percentual de 14% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 86% concordaram e a resposta média foi de 4,1. Pode-se concluir então que a maioria das instituições bancárias acham que os algoritmos de *machine learning* podem auxiliar seus analistas nas análises e concessões de crédito, dando-lhes segurança e suporte durante a tomada de decisão.



Figura 34 – Para gerentes e técnico, ML auxilia as análises e concessões de crédito?

Fonte: Elaboração própria

Nesta questão da Figura 34, os índices de concordância foram excelentes em ambos tipos de cargo. Com 88% dentre os técnicos e 80% dentre os gerentes. Diante destes resultados, afirma-se que as instituições bancárias como toda reconhecem a ajuda que *machine learning* pode trazer na análise e na concessão de crédito.

Atualmente, deparamo-nos com taxas de incumprimento cada vez mais elevadas. Então, resolveu-se neste momento analisar a Figura 35, e verificar se as instituições bancárias acham que com a implantação de *machine learning* no setor de análise crédito é possível melhorar os índices de incumprimento dos clientes da empresa.

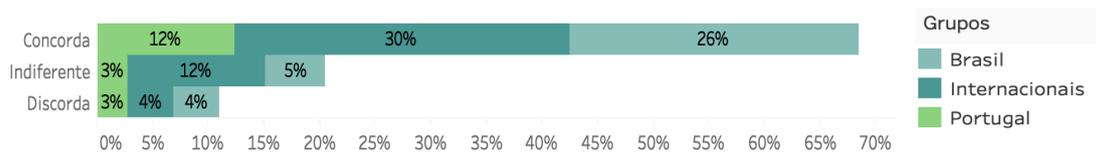


Figura 35 – ML melhora os índices de incumprimento dos clientes?

Fonte: Elaboração própria.

Dentre os entrevistados, cerca de 11% responderam em discordância desta questão, e 20% foram “indiferentes”, gerando um percentual parcial de 31% de percepção indiferente ou negativa a questão apresentada pela Figura35. Já nas respostas positivas, aproximadamente 69% de pessoas concordaram e a resposta média foi de 3,9. Logo, a maioria das instituições bancárias acreditam que a implantação *machine learning* no setor de crédito vai melhorar os índices de adimplência dos seus clientes.



Figura 36 – Para gerentes e técnicos, ML melhora os índices de incumprimento?

Fonte: Elaboração própria.

Dentre os entrevistados, gerou-se uma percentual concordância de 71,7 % dentre os técnicos e 60 % dentre os gerentes. Os gráficos da Figura 36, mostraram também que os percentuais de indiferença para esta questão foram um tanto altos, com 20% dos técnicos e 25% dos gerentes. Então, é na identificação destes funcionários que as instituições bancárias podem investir em capacitação de *machine learning* e assim diminuir os índices de rejeição em futuros projetos com esta tecnologia.

Devido à falta de garantias reais, os bancos necessitam obter mais informações dos clientes que possam servir de subsídios para a redução do risco.

Na Figura 37, analisou-se uma percepção sobre o uso de *machine learning* na análise qualitativa do crédito:

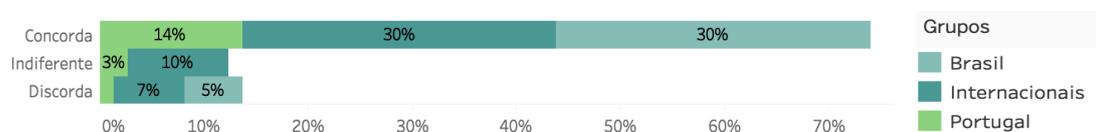


Figura 37 – ML auxilia na análise qualitativa dos C's do crédito

Fonte: Elaboração própria.

Nesta questão da Figura 37, obteve-se 13,7% dos entrevistados que discordaram, 13% foram “indiferentes”, gerando um percentual parcial de 26,7% de percepção indiferente ou negativa a questão. Além disso, foram 73,6% percepções positivas e a resposta média foi de 3,9. Isto posto, a maioria dos bancos acreditam que *machine learning* possa sim auxiliar na análise qualitativa de seus clientes.



Figura 38 – Machine Learning na análise qualitativa do crédito por tipo de cargo.

Fonte: Elaboração própria.

Já a questão da Figura 38, obteve-se também em sua maioria para concordância que *machine learning* pode auxiliar na análise qualitativa do crédito. Sendo que 81,13% dentre os técnicos e 55 % dentre os gerentes. O gráfico a nível de gerência também mostrou um valor alto de indiferença (25 %). Portanto, para a questão de *machine learning* no auxílio da análise qualitativa, cabe aos dirigentes das instituições bancárias, caso pretendam investir neste tema, definir estratégias junto aos seus gerentes.

Na Figura 39, pode-se analisar se *machine learning* pode melhorar e controlar rácios de liquidez, solvabilidade e rendibilidade:

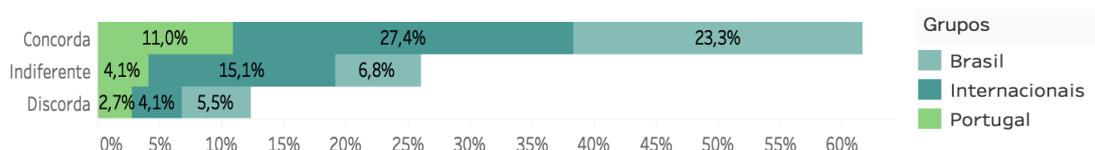


Figura 39 – Os RL, solvabilidade e rendibilidade são melhorados por ML?

Fonte: Elaboração própria.

De acordo com as respostas, 12% responderam em discordância e 26% foram indiferentes, gerando um percentual de 38% de percepção indiferente ou negativa a questão. Já nas respostas positivas, 62% concordaram com a questão da Figura 39. A resposta média foi de 3,8. Apesar do alto percentual de respostas como indiferentes, infere-se que a maioria das empresas acreditam que *machine learning* trará melhoras e controle aos rácios de liquidez, solvabilidade e rendibilidade.

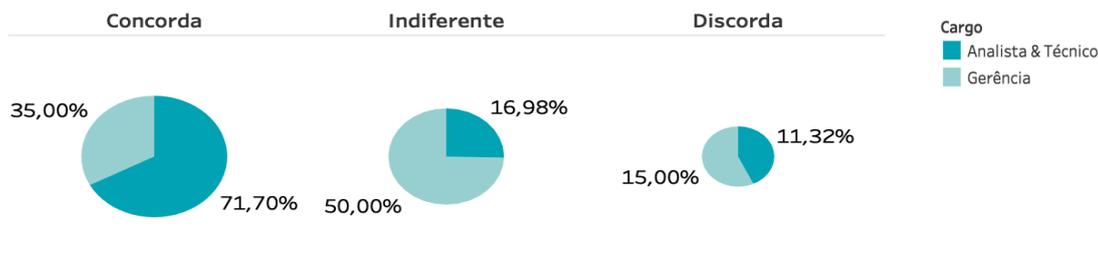


Figura 40 – ML nos RL, solvabilidade e rendibilidade por tipo de cargo.

Fonte: Elaboração própria.

Na relação de *machine learning* com melhoras nos rácios de liquidez, solvabilidade e rendibilidade, obteve-se uma discordância entre os níveis de cargo. Onde a Figura 40 demonstra que a maioria dos técnicos (71,7 %) concordam com a possibilidade de melhoria, enquanto que apenas 35% dos gerentes concordam. No detalhe da análise com os gerentes percebeu-se que a sua maioria (50%) foram indiferentes. Pode-se concluir que apesar da maioria dos entrevistados acreditarem que o uso de ML possa melhorar os rácios citados, os níveis de gerência ainda possuem um receio quanto a questão.

Para concluir, foram feitas três questões do tipo classificação, para analisar quanto a preferência e desafios das instituições bancárias em relação *machine learning* em risco de crédito, conforme ilustrado nas Figuras 41, 42 e 43.

	Cargo	
	Analista & Técnico	Gerência
Aprovação ou Rejeição	36 pts	13 pts
Acompanhamento	28 pts	13 pts
Cálculo de capital	13 pts	5 pts

Figura 41 – Em que áreas do crédito já se utiliza ou pretende utilizar ML?

Fonte: Elaboração própria.

Na Figura 41, percebeu-se uma preferência para o processo de aprovação e rejeição com 49 pts, seguido do processo de acompanhamento com 41 pts, e o processo cálculo de capital com apenas 18 pts. Cabe mencionar que como tratava-se de uma questão semiaberta, outras opções poderiam ser mencionadas pelos entrevistados. Foi possível observar as preferências por desenvolvimento de modelos de gestão de risco, na determinação do preço do risco, previsão de perda, monitorização de fraude financeira e

modelagem estatística. Quanto a preferência de processos de crédito para utilização de *machine learning* não se obteve diferenciação significativa entre os tipos de cargo. Ambos preferiram os processos de aprovação e acompanhamento como mais relevantes.

	Cargo	
	Analista & Técnico	Gerência
Pessoas	21 pts	13 pts
Tecnologia	27 pts	7 pts
Processos	22 pts	11 pts
Dados	19 pts	11 pts

Figura 42 – Quais são os maiores desafios para se implementar ML em RC?

Fonte: Elaboração própria.

Neste tópico exposto pela Figura 42, viu-se um equilíbrio nas respostas dos entrevistados com relação aos desafios para se implementar *machine learning* em risco de crédito. Como a diferença é mínima, considerou-se que os principais desafios são todos quatro itens do questionário original – “Pessoas (34 pts), Processos (34 pts), Tecnologia (33 pts) e Dados (30 pts)”. Como na questão anterior, esta também era do tipo semiaberta, o que se permitiu ver a escolha de um dos entrevistados por ver um desafio na fiabilidade da análise. Além disso, é interessante ressaltar a percepção dos gerentes que comparados aos outros desafios, não entendem a “Tecnologia” como um dos principais desafios. Os técnicos responderam que o maior desafio em se implementar *machine learning* no risco de crédito está relacionado a “Tecnologia”. Enquanto para os gerentes este seria o menor deles, destacando “Pessoas” como maior desafio para o uso de *machine learning* em risco de crédito.

Na Figura 43, de acordo com as respostas dos entrevistados, no risco de crédito, o algoritmo de árvore de decisão foi o preferido na escolha do que mais traz eficácia para os negócios do banco com o uso de *machine learning*. Sua pontuação foi de 193,9 pontos. Seguido do algoritmo de classificação do Naïve Bayes com 158,8 pontos e algoritmos de regressão linear, com 158,8 pontos.

Ainda nesta última análise da Figura 43, não se obteve diferença significativa entre os cargos. A diferença ocorreu com relação ao algoritmo de redes neurais que para os gerentes, estaria duas posições acima na classificação de preferências de eficácia.

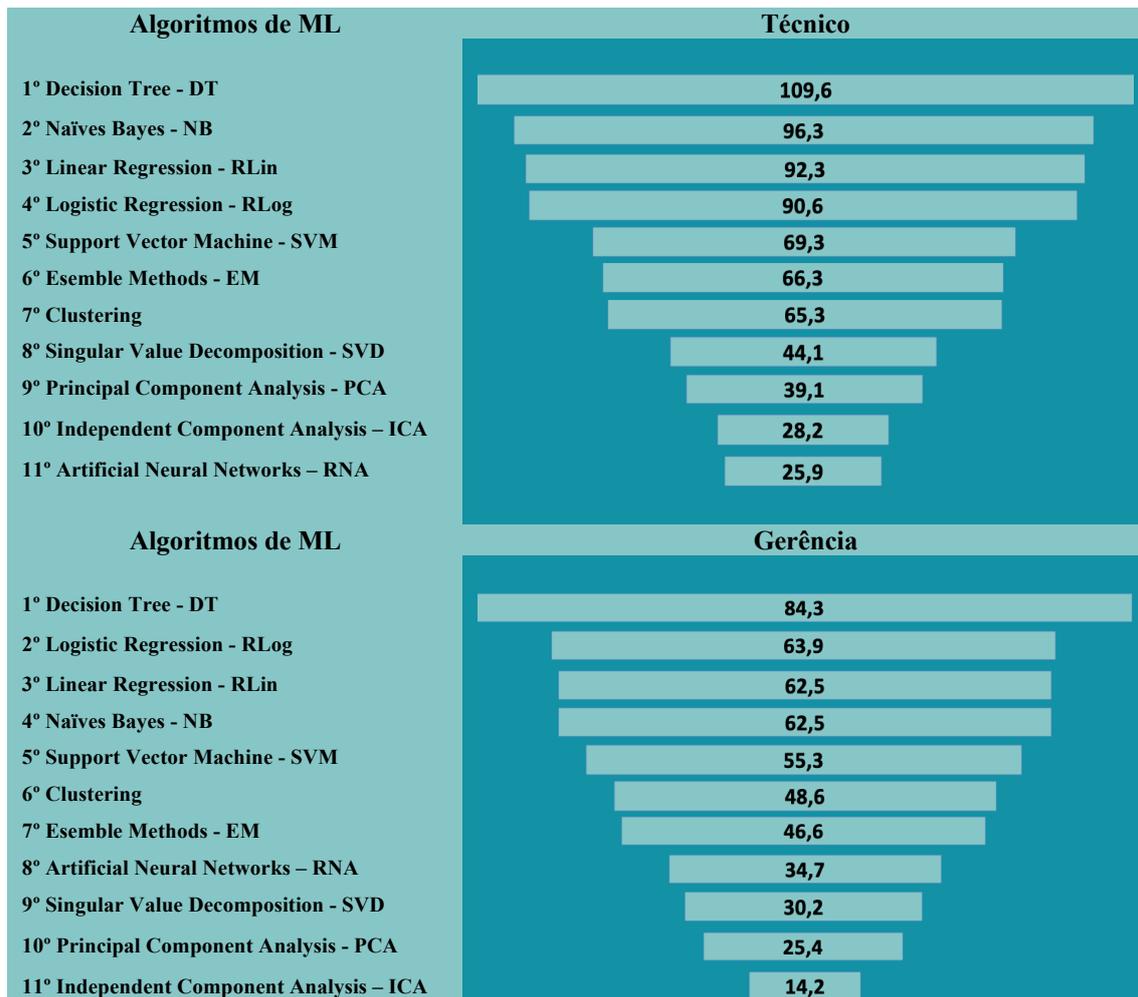


Figura 43 – ML trará mais eficácia para os negócios na ideia dos gerentes e técnicos.

Fonte: Elaboração própria.

Doravante, faz-se uma análise do questionário “Aos técnicos de crédito”, de forma integral das questões, através da pontuação de Likert. Esta análise está subdividida em duas seções. A primeira efetuará a análise da pontuação de Likert nas afirmações de *machine learning*. O objetivo desta primeira análise é verificar, de forma mais generalizada, como está o relacionamento das instituições com o assunto de *machine learning*.

A segunda será relativo as afirmações de *machine learning* na área de risco de crédito das instituições. O objetivo desta segunda é verificar o quanto que a área de crédito das instituições está envolvida com *machine learning*.

A fórmula utilizada foi:

$$P_{\max} = QML * CL_{\max} * E$$

onde: P_{max} = Pontuação máxima para as questões de ML;

QML = Número questões de ML;

CLmax = Maior valor da escala Likert;

E = Quantidade de entrevistados.

A. Afirmações de *Machine Learning*

Esta parte do questionário tem 9 questões de tipo Likert de 5 opções (-2 a 2). Aplicando a fórmula:

$$P_{\max} = 9 * 2 * 73 = 1314.$$

A partir do Pmax obteve-se a seguinte grelha de pontuação:

- 0 a 328, muito fraca a relação das instituições bancárias ML;
- 329 a 657, as IB's ainda têm muito a melhorarem com relação a ML;
- 658 a 986, as IB's têm muito mais pontos positivos do que negativos e com grande potencial para melhorias;
- 987 a 1314, as IB's já estão num estado de excelência com relação a ML.

Tabela 12 – Análise de Likert para as afirmações de ML

Questões	Pontuação	Acumulado
09	96	96
10	77	173
11	36	208
12	91	299
13	40	339
14	47	386
15	37	423
16	48	471
17	4	475
Total		475

Fonte: Elaboração própria.

Concluiu-se assim, que as afirmações relacionadas a *machine learning* acumularam um total de 475. O que pode inferir que as instituições bancárias ainda têm muito a melhorarem com relação a *machine learning*.

B. Afirmações de *machine learning* para risco de crédito

Esta parte do questionário tem 5 questões de tipo Likert de 5 opções (-2 a 2). Aplicando a fórmula:

$$P_{\max} = 5 * 2 * 73 = 730.$$

A partir do Pmax obteve-se a seguinte grelha de pontuação:

- 0 a 182, muito fraca a relação das instituições bancárias com ML em RC;
- 183 a 365, as IB's ainda têm muito a melhorarem com relação a ML em RC;

- 366 a 548, as IB's têm muito mais pontos positivos do que negativos e com grande potencial para melhorias;
- 549 a 730, as IB's já estão num estado de excelência com relação a ML em RC.

Tabela 13 – Análise de Likert para as afirmações de ML para risco de crédito

Questões	Pontuação	Acumulado
19	45	45
22	84	129
23	60	189
24	66	255
25	50	305
Total	305	

Fonte: Elaboração própria.

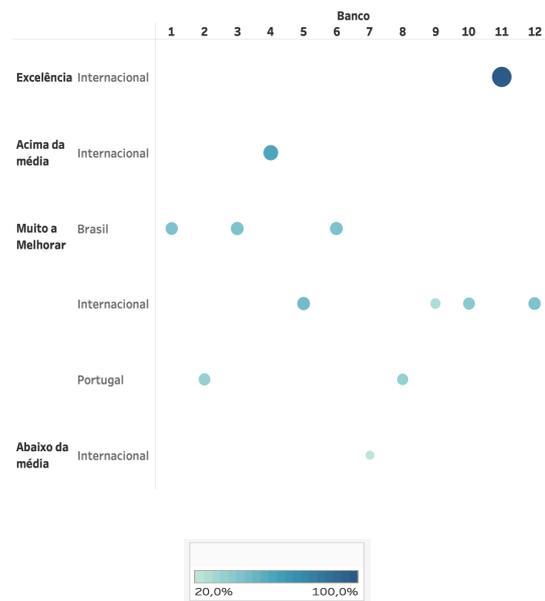
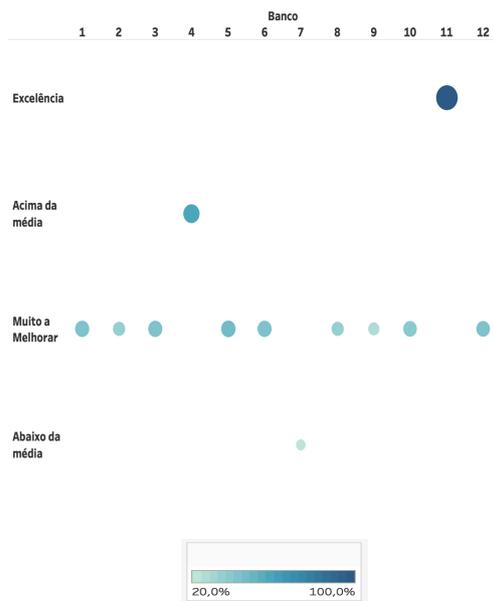


Figura 44 – Likert - ML para RC por banco.

Figura 45 – Likert - ML para RC por nicho.

Fonte: Elaboração própria.

Fonte: Elaboração própria.

Enfim, após a análise das questões supracitadas, pode-se concluir que *machine learning* em risco de crédito acumularam um total de 305. O que pode inferir que no geral as instituições bancárias ainda têm muito a melhorarem com relação a *machine learning* em risco de crédito. Além disso, as Figuras 44 e 45 trazem um detalhamento por banco e nichos respectivamente, o que possibilita descrever que na amostra, o banco 11, do nicho das instituições internacionais, já se encontra em excelência no tema. Seguido a este tem o banco 4 (nicho internacional) que possui um conceito acima da média. Destaca-se ainda que por exceto ao banco 7 (nicho internacional) que está muito abaixo da média, os demais bancos estão no patamar de que tem muito a melhorar na relação de ML em RC.

4.3. Questionário – Aos Executivos e Diretores de Crédito

4.3.1. Recolha de dados

Relativamente a recolha de dados do questionário “Aos Executivos e Diretores de Crédito”, a amostra consiste em 6 respostas de executivos ou diretores de créditos selecionados das instituições bancárias citadas no capítulo anterior. E este foi elaborado com o objetivo de validar o primeiro questionário no contexto do *machine learning* em risco de crédito de instituições bancárias. Com relação aos resultados foram considerados: resultados acima de 80% seriam ótimos; 50% a 80% seriam regulares e abaixo de 50% não desejados.

4.3.2. Resultados

Iniciou-se esta secção de resultados com a validação do porte das instituições bancárias inquiridas no questionário. Para este levantamento foi feita a Figura 46. Por questões de anonimato e sigilo acordado com os entrevistados, os bancos não foram identificados.



Figura 46 – Validação do perfil das IB's pelos executivos ou diretores de crédito.

Fonte: Elaboração própria.

A Figura 46, apresentou alguns indicadores com relação ao porte da empresa em estudo. Os resultados foram ótimos em 100% para os três indicadores. O que valida a amostra quanto ao seu porte institucional.

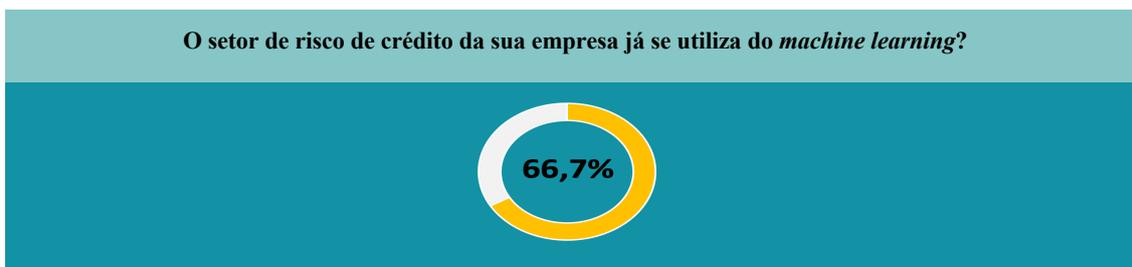


Figura 47 – Validação da utilização de Machine Learning em risco de crédito.

Fonte: Elaboração própria.

De acordo com as respostas dos executivos e diretores de créditos expostas na Figura 47, 66,7 % das instituições inquiridas já se utilizam de *machine learning* em risco de crédito. Portanto, esta questão também está validada com conceito regular para o estudo.

A seguir aprofundou-se, através das Figuras 48 e 49 na análise da atuação de ML no RC, conforme respostas obtidas no questionário “Aos executivos e diretores de crédito” de validação.

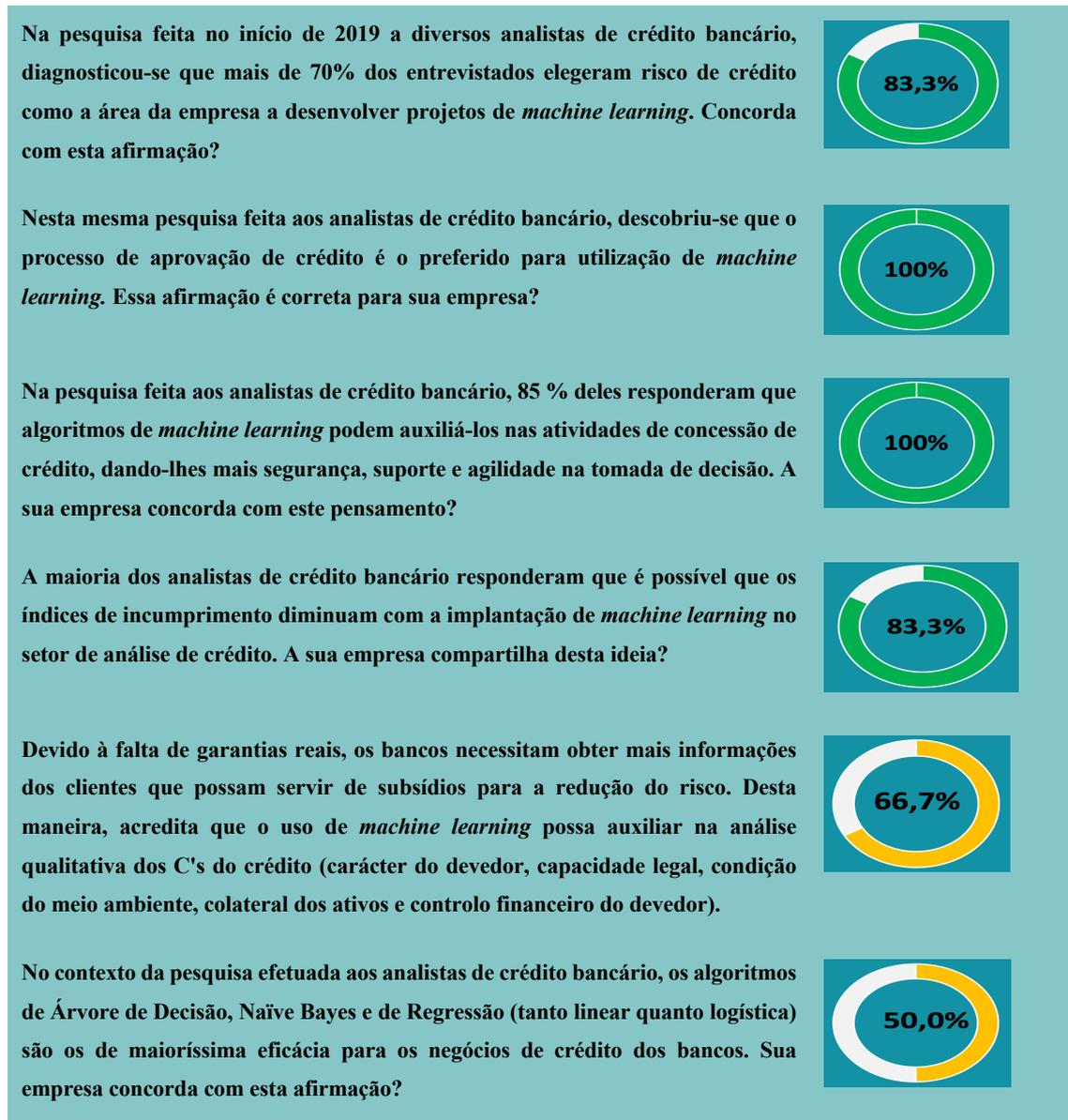


Figura 48 – Validação das questões de Machine Learning em risco de crédito.

Fonte: Elaboração própria.

Os resultados das principais questões do questionário 2 estão representados na Figura 48. A primeira questão foi validada pelos executivos ou diretores com resultado ótimo para a preferência de 70% dos analistas pelo uso de ML na área de crédito.

Também se validou com resultado ótimo para o processo de aprovação de crédito e que ML trará mais segurança, suporte e agilidade na tomada de decisão das concessões de crédito. Outro ponto que obteve validação ótima foi para ideia da maioria de analistas que é possível diminuir os índices de incumprimento através de ML. Entretanto, para a questão de ML auxiliar na análise qualitativa dos clientes, o conceito foi considerado aprovado com 66,7 % dos inquiridos.

E por último, a análise da Figura 48 validou regularmente para os algoritmos DT, NB, Rlin e Rlog como os mais eficazes para negócios de crédito. Neste item, os executivos ou diretores mencionaram que RNA e GA teriam eficácia melhor.

Para concluir a análise do questionário 2, foi elaborada a Figura 49 com os principais desafios elencados pelos executivos ou diretores das instituições bancárias.

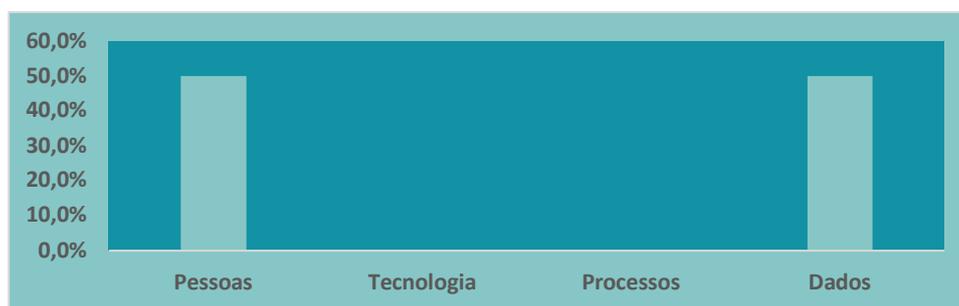


Figura 49 – Desafios de ML em risco de crédito na visão dos executivos ou diretores.

Fonte: Elaboração própria.

Conforme a Figura 49, os desafios selecionados pelos executivos e diretores de crédito foram apenas dois: Pessoas e Dados.

4.4. Discussão

Neste item, apresentou-se as interpretações e ilações auferidas dos dados dos questionários feitos por este estudo. Assim sendo, organizou-se este item por subtítulos relacionados às questões dos questionários. O primeiro subtítulo fez um breve relato das informações do próprio questionário. No segundo analisou-se o perfil do entrevistado, dando destaque a idade, tempo de empresa, formação acadêmica, cargo e se já participa de alguma atividade com *machine learning*. Com intuito de selecionar empresas de mesmo porte o terceiro subtítulo tratou-se do perfil das instituições bancárias dos entrevistados, face a quantidade de funcionários, volume de negócios e a correspondência deste negócio em crédito. O quarto perguntou-se aos especialistas em crédito a respeito do conhecimento de *machine learning* dentro da instituição que trabalha como todo.

Por último, fez-se o quinto subtítulo para identificar o uso e o conhecimento do que *machine learning* pode proporcionar aos processos de risco de crédito.

4.4.1. Questionário

Nesta pesquisa foram feitos dois questionários: um aos técnicos de crédito das instituições bancárias da amostra. O segundo foi um refinamento das respostas do primeiro questionário, destinados aos executivos e diretores, no intuito de validação.

A percentagem de respostas foi de 14% (73 respostas) para o questionário “Aos técnicos de crédito” e 50% (6 respostas) para o questionário “Aos executivos e diretores de crédito”. Acredita-se também que aplicações periódicas deste questionário, juntamente com a divulgação pública dos resultados e medidas de conscientização quanto ao objetivo de melhorar todo o sistema de crédito bancário através da utilização massiva da tecnologia de *machine learning*, possam, no futuro, contribuir com o aumento da percentagem de resposta.

Por fim, percebeu-se que o desempenho dos inquiridos foi muito bom. Sendo gasto, pela maioria, até dez minutos para responder o questionário.

4.4.2. Perfil do Entrevistado

Com intuito de compreender o perfil dos entrevistados foram feitas cinco questões no questionário “Aos técnicos de crédito”.

A primeira relacionou-se com o cargo. Onde percebeu-se que a maioria dos pesquisados eram analistas ou técnicos de crédito. Apenas 27% exerciam uma função de coordenação ou gerência. Com isso, pode-se inferir, conforme a amostra, que o perfil de técnicos e analistas são os que mais participam dos inquéritos acadêmicos de *machine learning* para risco de crédito.

A seguir foi a questão da idade. Que de acordo com as respostas dos inquiridos, a faixa etária que possuiu maior quantidade de participação foi a “de 25 a 35 anos”. Fez-se uma correlação com cargo, onde o resultado foi que em todas as faixas etárias da amostra possuem mais analistas e técnico do que a função de gerência. Exceto a faixa de “45 a 60 anos” que tiveram números bem próximos.

Depois disso, fez-se a questão sobre a formação acadêmica do entrevistado. Nesta concluiu-se que a maioria da amostra possui um título de mestrado. Se adicionar a esta

maioria, o conjunto daqueles possuem uma especialização, alcança quase 80% do total dos entrevistados. Isso leva a interpretar que amostra tem um ótimo nível intelectual.

No tocante ao tempo de trabalho na instituição bancária descobriu-se que a amostra além de ter um nível académico ótimo, tem também um nível ótimo de experiência em instituição bancária. Mais de 50% dos entrevistados possuem mais de 5 anos de empresa e a faixa com maior percentual foi a de “mais de 10 anos” de empresa (36%).

Para finalizar o perfil do entrevistado, questionou-se os entrevistados se trabalham com o tema deste estudo. Para esta questão, inferiu-se que a metade da amostra já trabalha com *machine learning* ou tecnologias afins.

Em síntese, o perfil do entrevistado é técnico, de média idade, bom nível académico, experiente em instituições bancárias e que a metade já trabalha com *machine learning*.

4.4.3. Perfil das Instituições Bancárias

No seguimento a discussão dos resultados, tratou-se das informações das instituições bancárias onde os entrevistados trabalham atualmente.

Foram expostos por mais de 80% dos entrevistados que suas empresas possuem mais de 1000 funcionários.

Face ao volume de negócio, percebeu-se que 97% das instituições da amostra estão no patamar “acima de 50 milhões de Euros”. Ao detalhar esta questão, descobriu-se que apenas um banco português e outro do nicho internacionais tiveram índices abaixo dos 100% (89% e 91% respetivamente).

Uma vez que o volume de negócio foi aprovado, partiu-se para análise do percentual em crédito. Neste ponto, 85% dos entrevistados responderam que suas empresas são do porte comercial em crédito de mais de 50% de seus negócios.

Em suma, as instituições bancárias da amostra são, em sua maioria, grandes, com alto volume de negócios e com grande percentual em crédito. Cabe salientar que 100% dos diretores e executivos inquiridos validaram esta informação.

4.4.4. Conhecimento em *Machine Learning*

Este item traz as interpretações a respeito das percepções dos entrevistados em relação a *machine learning*. Foram feitas 10 questões no questionário “Aos técnicos de crédito”.

Por este motivo, elaborou-se a Tabela 14 com as 9 primeiras questões e os percentuais de concordância dos entrevistados:

Tabela 14 – Percentual de concordância nas afirmações de ML

Questões	% Concordância
ML trará benefícios significantes?	90%
ML estará em todos produtos e serviços?	82%
Sua empresa atua com ML?	49,4%
Sua empresa pensa investir em ML até 2020?	79%
Falta profissionais qualificados em ML?	55%
Executivos compreendem os benefícios de ML?	17,8%
Executivos acreditam que ML é um processo longo e lento?	58,9%
Precisam melhorar os dados para depois usar ML?	60,3%
Terceirização dos serviços de TI para ML?	42,5%

Fonte: Elaboração própria.

De acordo com os resultados referenciados na Tabela 14, deduziu-se que os entrevistados acreditam que ML trará benefícios as instituições bancárias e que no futuro estará em todos os produtos e serviços da empresa. Além disso, constatou-se que a metade das instituições ainda não atuam efetivamente com ML. Por outro lado, acreditam que até 2020 pensam em investir em projetos de ML.

Todavia, percebeu-se que há uma divisão com relação a falta de profissionais qualificados. Um pouco mais da metade respondeu em concordância com esta falta de profissionais.

Outro ponto a ressaltar, foi que os entrevistados acreditam que seus executivos não compreendem os benefícios de ML, e que eles consideram o desenvolvimento em ML como um processo longo e lento.

Já em relação da necessidade primária de melhorar os processos a nível de dados para então investir em ML, o resultado foi que a maioria acredita nesta ideia.

A maioria dos entrevistados também reconhece que a terceirização em TI para ML não seja um fator primordial a evolução dos serviços da empresa.

Em continuidade a análise do conhecimento de ML, fez-se a décima questão. Esta questionou-se sobre qual área da empresa seria a mais adequada para o desenvolvimento de um projeto de ML. De acordo com os entrevistados, a área de detecção de fraude seria a área da empresa mais adequada para uso de ML. Com um resultado bem próximo, a segunda área mais referenciada foi a de risco de crédito.

Para sumarizar o conhecimento de ML, utilizou-se o cálculo do item 4.2.2.A. Este avaliou os resultados de todas as questões citadas acima de forma generalizada. Alcançando o valor de 475 pontos. O que corresponde a inferir que as instituições bancárias como toda ainda têm muito a melhorarem com relação a machine learning.

4.4.5. Conhecimento em *Machine Learning* para Risco de Crédito

Do conhecimento de ML em RC, foram elaboradas 8 questões no questionário “Aos técnicos de crédito”. E suas respostas foram validados em 6 questões do questionário “Aos executivos e diretores de crédito”.

Para melhor explanação das interpretações, utilizou-se a mesma estratégia do item anterior, através da Tabela 15 apresentou-se as cinco primeiras questões.

Tabela 15 – Percentual de concordância nas afirmações de ML para RC

Questões	% Concordância
O RC da sua empresa já atua ou pensa em atuar com ML?	65,8%
ML pode auxiliar nas análises e concessões de crédito?	86%
ML melhora os índices de incumprimento?	69%
ML auxilia a análise qualitativa do crédito?	73,6%
ML controla a liquidez, solvabilidade e rendibilidade?	62%

Fonte: Elaboração própria.

Dos resultados referenciados na Tabela 15, inferiu-se que mais da metade das instituições bancárias já atuam ou pensam em atuar com ML. Além disso, identificou-se através das respostas da maioria dos entrevistados que o uso de ML pode auxiliar as análises e concessões de crédito; melhorará os índices de incumprimento; auxiliará a análise qualitativa do crédito. Bem como, trazer melhoras e controle aos rácios de liquidez, solvabilidade e rendibilidade das instituições bancárias.

Em continuidade a análise do uso de ML em RC, fez-se a sexta questão em face da área do crédito que deve usar, primeiramente, ML. Cabe mencionar que as três principais áreas (aprovação, acompanhamento e cálculo de capital) podem se beneficiar das vantagens de ML. Porém, a área com maior pontuação e validada pelo executivos e diretores selecionados para este estudo foi a área de aprovação de crédito.

A sétima questão tratou-se dos principais desafios na visão dos inquiridos. Desse modo, por equilíbrio às respostas do questionário “Aos técnicos de crédito”, elencou-se quatro principais desafios: Pessoas (34 pts), Processos (34 pts), Tecnologia (33 pts) e Dados (30 pts). Por outro lado, os executivos e diretores, de crédito, envolvidos neste estudo, validaram como desafio, apenas: Pessoas e Dados.

Por último, os inquiridos foram questionados a respeito de quais algoritmos de ML trarão mais eficácia para os negócios das instituições bancárias. O algoritmo que teve unanimidade dentre todos os participantes e validado por 50% dos executivos e diretores de crédito deste estudo, foi: *decision tree* (193,9 pts).

Cabe salientar que a outra metade dos executivos e diretores de crédito deste estudo preferiram aos algoritmos de: *neural network*; *genetic algorithm*.

Ainda em relação a esta última questão, com as respostas dos inquiridos formou-se um ranking dos algoritmos, descrito na Tabela 16.

Tabela 16 – Ranking dos algoritmos de ML - Questionário

Ranking	Algoritmo ML	Pontuação
1º	Decision Tree - DT	193,9
2º	Naïves Bayes - NB	158,8
3º	Linear Regression - RLin	154,8
4º	Logistic Regression - RLog	154,5
5º	Support Vector Machine - SVM	124,6
6º	Esemble Methods - EM	113,9
7º	Clustering	112,9
8º	Singular Value Decomposition - SVD	74,3
9ª	Principal Component Analysis - PCA	64,5
10º	Artificial Neural Networks - RNA	60,6
11º	Independent Component Analysis - ICA	42,4

Fonte: Elaboração própria

Após a formação deste ranking, o estudo ficou com quatro visões diferentes em relação aos algoritmos: A visão face a literatura descrita pela Tabela 9, a visão do autor (R. A. Silva, 2014), visão do questionário “Aos técnicos de crédito” e “Aos executivos e diretores de crédito”. Para fins deste estudo elencou-se como principais algoritmos, por aparecerem mais vezes com as melhores classificações, Rlog, DT, RNA.

Da mesma forma do item anterior, a síntese da percepção do conhecimento em ML para RC foi feita através de um cálculo hipotético sob as questões envolvidas. Utilizou-se o cálculo do item 4.2.2.B. Este avaliou os resultados de todas as questões citadas acima de forma generalizada. Alcançando o valor de 305 pontos. O que corresponde a inferir que as instituições bancárias ainda têm muito a melhorarem com relação a *machine learning* em risco de crédito.

Capítulo 5 – Conclusão

Quando se iniciou este trabalho de pesquisa, constatou-se que devido as constantes mudanças no sistema económico mundial, principalmente após as recentes crises financeiras, as instituições bancárias necessitavam rever seus processos de concessão de crédito para torná-los mais ágeis, com mais critérios e rigor. Que os processos permitissem a previsão do comportamento dos clientes para melhor adequação do crédito. Por isso era importante estudar sobre ferramentas de tecnologia, como *machine learning* para risco de crédito.

Diante disso, a pesquisa teve como principal objetivo de fazer uma revisão na literatura sobre a utilização bancária da tecnologia de *machine learning*, em especial para a área de crédito. Constata-se que o objetivo geral foi atendido porque efetivamente o trabalho conseguiu demonstrar de várias formas que *machine learning* pode ser a ferramenta de transformação positiva para os processos de concessão de crédito bancário.

O primeiro objetivo específico era descobrir qual seria o valor de *machine learning* para o risco de crédito. Entende-se que este foi atendido porque descobriu-se tanto na literatura, mas principalmente nas respostas dos questionários aos especialistas de crédito bancário, que *machine learning* pode:

- Auxiliar as atividades de concessão de crédito, promovendo mais agilidade, suporte e segurança na tomada de decisão;
- Reduzir os índices de incumprimento. Uma vez que o uso de *machine learning* melhora a análise de classificação do cliente, torna a predição mais assertiva, consegue extrair informações estruturadas de dados não estruturados, otimiza os parâmetros de regressão, diminui o erro, enfim, calcula o melhor resultado;
- Mitigar os riscos. Visto que *machine learning* traz eficácia e eficiência para a análise qualitativa do cliente quanto: caráter do devedor; capacidade legal; condições de meio ambiente, colateral dos ativos e melhor controlo do cliente.

O segundo objetivo específico do estudo era selecionar os fatores de sucesso de uma implementação de *machine learning* em risco de crédito. Esta meta também foi alcançada pelo motivo que se constatou na literatura diversas utilidades de *machine learning* para risco de crédito, existência de diversos algoritmos (modelos). Cabe a instituição financeira bancária eleger, através do seu perfil, estratégias e testes, qual o seu modelo campeão.

Já o terceiro objetivo específico era identificar os desafios para uma implementação de *machine learning* para risco de crédito. Isso foi conquistado principalmente através das respostas dadas pelos analistas de crédito inquiridos nos questionários. Onde não se destacou um maior. Foram quatro principais desafios com percentuais bem próximos uns dos outros: dados, pessoas, processos e tecnologia.

No escopo da nossa investigação foi direcionado a compreensão do uso de *machine learning* no processo de risco de crédito das instituições bancárias pré-selecionadas, através de questionários. Os questionários foram respondidos, durante os meses de janeiro a março de 2019, por uma amostra de analistas de crédito de 12 instituições bancárias: três do Brasil, duas de Portugal, duas dos EUA, duas da China, uma de Espanha, uma do Reino Unido e uma de França. Estas respostas foram validadas, através de um segundo questionário, por um grupo de executivos e diretores de créditos das mesmas instituições citadas acima. Estes questionários nos forneceram informações que nos permitem acreditar que as instituições bancárias já saíram do estágio inicial e estão num segundo estágio de quatro possíveis para chegarem a excelência no assunto. Este segundo estágio reflete dizer que, de acordo com a amostra, instituições bancárias já saíram do âmbito do estudo, já possuem pequenas implementações, porém ainda têm muito a melhorarem com relação a *machine learning* em risco de crédito.

Em suma, baseado na revisão da literatura e nas análises dos questionários, foi possível concluir que o método de análise de dados, através da tecnologia de *machine learning* trará mais eficiência a diversos processos de risco de crédito. Isso porque constatou-se que *machine learning* melhora a assertividade da classificação, padroniza comportamentos, minimiza o risco de falhas, ou seja, proporciona a análise de um volume maior de informações, mais agilidade e qualidade aos processos de risco de crédito.

Por outro lado, em momentos de instabilidade financeira, instituições de alto risco, e uma vez que a utilização de *machine learning* ainda é emergente, onde um erro de análise pode causar um prejuízo grande, ou até mesmo a falência, pode-se dizer que este estudo apresentou um resultado dentro do esperado de que os bancos ainda estão cautelosos em sua maioria. Porém, se almejam vantagem competitiva, precisam avançar mais na relação de *machine learning* em risco de crédito.

Diante da metodologia usada observa-se que o estudo poderia ser construído através de uma pesquisa de mercado mais ampla de instituições bancárias. Já que perante das

limitações financeiras, de tempo e geográficas, só foi possível analisar uma amostra pequena de 12 instituições.

Além disso, cabe mencionar como dificuldade a desconfiança que as pessoas de instituições bancárias têm em revelar informações para acadêmicos. Muitas deixaram de responder aos questionários por alegarem ser informações sigilosas ou estratégicas.

Uma outra limitação em decorrência do tempo, seria o aprofundamento do tema com um estudo de caso real, onde pudesse comparar um processo de risco de crédito antes e depois de *machine learning*.

Como mencionado anteriormente, a quantidade foi limitada. Em função disso, recomenda-se para trabalhos futuros a incorporações de novas instituições bancárias.

Bibliografia

Abellán, J., & Mantas, C. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 41, 3825–3830.

ABI Research. (2018). *Global connectivity will remain reliant on bluetooth, Wi-Fi and RFID, Even as LPWA catches up*. Consultado em 2018, Dezembro. Disponível em <https://www.abiresearch.com/press/global-connectivity-will-remain-reliant-bluetooth-wi-fi-and-rfid-even-lpwa-catches/>

Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. *SMU HACNet Lab Southern Methodist University Dallas, TX 75275*.

Adamko, P., Klieštík, T., & Mišanková, M. (2014). Applied comparison of selected credit risk models. *ResearchGate*.

Agrawal, D. (2014). Analytics based decision making.

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, 222, 168–178.

Ala'raj, M., & Abbod, M. (2016). Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems* 104, 89–105.

Aleskerov, E., Freisleben, B. & Rao, B. (1997). Cardwatch: a neural network based database mining system for credit card fraud detection. In *Proceedings of Computational Intelligence for Financial Engineering*, 173-200.

Ali, A., & Daly, K. (2010). Macroeconomic determinants of credit risk: Recent evidence from a cross country study. *International Review of Financial Analysis*, 19(3), 165-171.

Al-Jarrah, O., Yoo, P., Muhaidat, S., Karagiannidis, G., & Taha, K. (2015). Efficient machine learning for big data: A Review. *Big Data Research*, 2(3), 87–93.

Altman, E., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21, 1721–1742.

Amanze, B., Asogwa, D., & Chukwuneke, C. (2019). Enhanced credit worthiness of bank customer in Nigeria using machine learning and digital nervous system. *International Educational Journal of Science and Engineering (IEJSE)*, 2(1), Jan-Feb 2019 | E-ISSN N°: 2581-6195.

Amaral, A., Casanova, C., Cruz, J., Simões, L., & Antunes, P. (1997). *Financiamento e Crédito Bancário* (2ª ed.). Lisboa: Instituto Superior de Gestão Bancária.

- Amaral, M. P. (2015). *Tipos de riscos na actividade bancária*. Consultado em 2018, Dezembro. Disponível em <http://www.oroc.pt/fotos/editor2/Revista/69/Gestao.pdf>
- Andrade, F. (2004). Desenvolvimento de modelo de risco de portfólio para carteiras de crédito a pessoas físicas. São Paulo, 2004. Tese (Doutorado em Administração) – *Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas*.
- Angelini, E., Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48 - Elsevier, 733–755.
- APB. (2015). *Associação Portuguesa de Bancos*. Consultado em 2018, Dezembro. Disponível em <http://www.apb.pt>
- Aragão, C. S., Carvalho, L., & Barros, M. (2002). Análise do risco de uma carteira de crédito por meio de simulação de Monte Carlo. *Resenha BM&F*, 152, 49-56.
- Arewa, A., Nwachukwu, U., & Owoputi, J. (2013). Bank credit risk and interest rate volatility-granger causality vs. var-garch approach international. *Journal of Business and Management Review*, 1(2), 26–34.
- Arminger, G., Enache, D., & Bonne, T. (1997). Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis, and feedfor-ward neural networks. *Computational Statistics*, 12, 293-310.
- Arora, R., Fan, C., & Leblanc, G. (2019). Liquidity management of Canadian corporate bond mutual funds: A machine learning approach. *Staff Analytical Note*.
- Arraes, D., Semolini, R., & Picinini, R. (1999). Arquiteturas de redes neurais aplicadas a data mining no mercado financeiro: uma aplicação para a geração de credit ratings. In: *Congresso Brasileiro de Redes Neurais*, 4, 1999, São José dos Campos: Sociedade Brasileira de Redes Neurais.
- Asha, K. A., Patyal, A., & Chen, H.-M. (2018). Generation of PUF-Keys on FPGAs by K-means Frequency Clustering. *IEEE*.
- Aver, B. (2008). An empirical analysis of credit risk factors of the Slovenian banking system. *Managing Global Transitions*, 6, 317–334.
- Bücher, M., Kampen, M., & Krämer, W. (2013). Reject inference in consumer credit scoring with nonignorable missing data. *Journal of Banking & Finance*, 37, 1040-1045.
- Balaniuk, R. e. (2013). Corruption risk analysis using semi-supervised Naïve Bayes classifiers. *International Journal of Reasoning-based Intelligent Systems*, 5(4), 237-245.
- Balbi, S. (2006). *Banco Fica com Parcela Maior do “Spread”*. Consultado em 2019, Março. Disponível em www1.folha.uol.com.br/folha/dinheiro/ult91u105585.shtml
- Banco de Portugal. (2007). *ConsultasPublicas*. Consultado em 2018, Dezembro. Disponível em http://www.bportugal.pt/pt-PT/OBancoeoEurosistema/ConsultasPublicas/Documents/Consulta_BP_2_07_M A R.pdf

Banco de Portugal. (2015). Glossário do Banco de Portugal de 10 de dezembro de 2015.

Banco do Brasil. (2018). Relações com Investidores: *Tipos de riscos*. Consultado em 2018, Dezembro.

Disponível em <https://www.bb.com.br/portallbb/page3,136,2545,0,0,1,8.bb>

Banco Mundial. (2014). *RDM - Relatório de Desenvolvimento Mundial de 2014 – Risco e Oportunidade: Gestão do Risco para o Desenvolvimento*. Consultado em 2018, Dezembro. Disponível em

http://siteresources.worldbank.org/EXTNWDR2013/Resources/8258024-1352909193861/89369351356011448215/8986901-1380730352432/POR_StandaloneOverview.pdf

Banco Popular do Brasil. *O que é crédito*. Consultado em 2018, Dezembro.

Disponível em

<https://www.bb.com.br/portallbb/page251,8900,8923,0,0,1,0.bb?codigoMenu=5415&codigoNoticia=8133>

Banco Santander Totta. (2015). *Relatório 1o Semestre de 2015*. Consultado em 2018, Dezembro. Disponível em

https://www.santandertotta.pt/pt_PT/pdf/RelGestaoBST1Sem2015.pdf

Bank of China. (2017). *Pillar 3 Disclosures 2017*. Consultado em 2018, Dezembro.

Disponível em

<http://pic.bankofchina.com/bocappd/uk/201803/P020180328577372288229.pdf>.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417.

Baum, L. (1972). An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov Process.

BCB. (2017). Resolução Nº 4.557 de 23 de fevereiro de 2017 - Banco Central do Brasil.

BCBS. (2000). *Bank for International Settlements. Principles for the management of credit risk*. Consultado em 2019, Janeiro.

Disponível em <https://www.bis.org/publ/bcbs75.htm>

Bekhet, H., & Eletter, S. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Science Direct Review of Development Finance*, 4, 20–28.

Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36, Elsevier - ScienceDirect, 3302–3308.

Bellotti, T., & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4), Elsevier, 563–574.

- Ben-David, A. (1995). Monotonicity maintenance in information-theoretic machine learning algorithms. *Machine Learning*, 19, Kluwer Academic Publishers, 29-43.
- Bequé, A., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*, 86, 42–53.
- Bertsekas, D. (2012). Dynamic programming and optimal control: Approximate dynamic programming. *Athena Scientific*, II.
- Bertsekas, D., & Tsitsiklis, J. (1996). Neuro-Dynamic Programming, *Athena Scientific*.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613.
- Bijak, K., & Thomas, L. (2012). Does segmentation always improve model performance in credit scoring? *Expert Systems with Applications*, 39, 2433-2442.
- Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*, 40, 356-364.
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235-249.
- Bottou, L., & Vapnik, V. (1992). Local Learning Algorithms. *Neural Computation*, 4(6), 888–900.
- Boyan, J. (2002). Technical update: Least-squares temporal difference learning. *Boston: Kluwer Academic Publishers*, 233-246.
- Bradtke, S., & Barto, A. (1996). Linear least-squares algorithms for temporal difference learning. *Kluwer Academic Publishers*.
- Brause, R., Langsdorf, T. & Hepp, M. (1999). Neural data mining for credit card fraud detection. Proceedings: In *11th IEEE International Conference on Tools with Artificial Intelligence*.
- Brito, G. A., & Assaf Neto, A. (2008). Modelo de classificação de risco de crédito em empresas. *Revista de Contabilidade e Finanças*, 19(46), 18-29.
- Brown, I., & Mues, C. (2012). An experimental comparison of classifications algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39, 3446-3453.
- Bucur, I., & Dragomirescu, S. (2014). The influence of macroeconomic conditions on credit risk: Case of Romanian banking system. *Studies and Scientific Researches, Economics Edition*, 19.

Budhathoki, D., Dasgupta, D., & Jain, P. (2018). Big data framework for finding patterns in multi-market trading data. In: Chin F., Chen C., Khan L., Lee K., Zhang L.J., Big Data 2018. *Lecture Notes in Computer Science*, 10968. Springer, Cham.

Caeiro, V. (2011). *Avaliação do risco de crédito de clientes empresariais. Levantamento de requisitos e estimação de modelos* (Dissertação de Mestrado em Finanças). Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Camarinha, M. (2009). Auditoria na Banca Utilizando Redes Bayesianas.

Caouette, J., Altman, E., Narayanan, P., & Nimmo, R. (2008). Managing credit risk: The great challenge for the global financial markets. *John Wiley & Sons*.

Carvalho, R. S., Carvalho, R. N., Ladeira, M., Monteiro, M. M., & Mendes, G. L. O. (2014). Using political party affiliation data to measure civil servants' risk of corruption. In: Intelligent Systems (BRACIS). *2014 Brazilian Conference on IEEE*, 166-171.

Carvalho, P. (2009). *Fundamentos da Gestão de Crédito: Uma contribuição para o valor das organizações*. Silabo, 1ª edição, Lisboa. ISBN: 978- 972-618-553-6.

Castro, V. (2013). Macroeconomic determinants of the credit risk in the banking system: The case of the GIPSI. *Elsevier, Economic Modeling*, 31, 672–683.

Chartier, B. & Spillane, T. (2000). Money laundering detection with a neural network. *Business Applications of Neural Networks*. Singapore: World Scientific, 159-172.

Chen, M. C., Huang, S. H., & Chen, C. M. (2002). Credit classification analysis through the genetic programming approach, Taipei: Proceedings of the 2002. *International Conference in Information Management*. Tamkang University.

Chen, W., Ma, C., & Ma, L. (2009). Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications* 36, 7611–7616.

Chen, Y.-S., & Cheng, C.-H. (2013). Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. *Knowledge-Based Systems, Elsevier*, 39, 224–239.

Cheng, X., Fang, L., Yang, L., & Cui, S. (2018). Mobile big data. *Wireless Networks*. Springer, Cham.

Chung, H., Ho, C., & Hsu, C. (2011). Support vector machines using Bayesian-based approach in the issue of unbalanced classifications. *Expert Systems with Applications*, 9(38), 11447–11452.

Citi bank. (2018). *Basel III Advanced Approaches Disclosures*. June 30,2018. Consultado em 2018, Dezembro. Disponível em <https://www.citigroup.com/citi/investor/data/b3ad180630.pdf?ieNocache=118>

Coelho, L., Raittz, R., & Trezub, M. (2006). FControl: sistema inteligente inovador para detecção de fraudes em operações de comércio eletrônico. *Gestão e Produção*, 13(1), 129-139. ISSN 0104-530X .

Collin-Dufresne, P., Goldstein, R., & Martin, J. (2001). The determinants of credit spread changes. *The Journal of Finance*, 56(6).

Cooper, P. (2019). *28 Twitter Statistics All Marketers Should Know in 2019*. Hootsuite Social Media Management. Consultado em 2019, Janeiro. Disponível em <https://blog.hootsuite.com/twitter-statistics/>

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 273–297.

COSO. (2007). Committee of Sponsoring Organizations of the Treadway Commission. *Understanding the new integrated ERM framework – Application Techniques*. Jersey City, NJ.

COSO. (2017). Committee of Sponsoring Organizations of the Treadway Commission. *Enterprise Risk Management – Integrating with Strategy and Performance*.

Crone, S., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, 28, 224-238.

Crook, J. N., Edelman, D. B., & Thomas, L. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447–1465.

Cubiles-De-La-Vega, M.-D., Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2013). Improving the management of microfinance institutions by using credit scoring models based on statistical learning techniques. *Expert Systems with Applications*, 40, 6910–6917.

Dantas, R., & Souza, S. (2008). Modelo de risco e decisão de crédito baseado em estrutura de capital com informação assimétrica. *Pesquisa operacional*, 28(2), 263-284.

Dean, J. (2014). *Big Data, Data Mining, and Machine Learning. Value Creation for Business Leaders and Practitioners*. John Wiley & Sons, Inc., Hoboken, New Jersey.

De-La-Veja, M. D., Blanco-Olivier, A., Pino-Mejias, R., & Lara-Rubio, J. (2013). Improving the management from microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert Systems with applicattions*, 40, 6910-6917.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing Big Data issues in scientific data infrastructure. Collaboration Technologies and Systems. doi: 10.1109/CTS.2013.6567203.

Dima, A. M., & Orzea, I. (2014). Risk Management in Banking. *AcademyPublish.org*.

Dinov, I. D. (2008). Expectation maximization and mixture modeling tutorial. *UCLA: Statistics Online Computational Resource*. Consultado em 2019, Fevereiro. Disponível em <https://escholarship.org/uc/item/1rb70972>

- Dorronsoro, J. R., Ginel, F., Sanchez, C. & Cruz, C. S. (1997). *Neural fraud detection in credit card operations*, 8, 827-834.
- Do, T.-N., & Poulet, F. (2015). Random local SVMs for classifying large datasets, in future data and Security Engineering SE. *Springer International Publishing, 9446*, 3–15.
- Doumpos, M., Zopounidis, C., & Adriosopoulos, K. (2014). Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from listed firms. *Journal of banking & finance*.
- Duan, J., & Shrestha, K. (2011). Statistical credit rating methods. *Global Credit Review*, 1(1), 43-64.
- EFQM. (2005). European Foundation for Quality Management. *EFQM Framework for Risk Management*. Brussels.
- Eifert, D. (2003). *Análise quantitativa na concessão de crédito versus inadimplência: um estudo empírico* (Dissertação de Mestrado em Administração). Universidade Federal do Rio Grande do Sul, Porto Alegre, 63-65.
- Elattar, E., Goulermas, J., & Wu, Q. (2010). Electric load forecasting based on locally weighted support vector regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(4), 438–447.
- Elliott, R., Siu, T., & Fung, E. (2014). A double HMM approach to Altman Z-scores and credit ratings. *Expert systems with applications*, 41, 1553-1560.
- Enciclopédia Britannica. (2019). Definição: Machine learning.
- Erdal, H., & Ekinçi, A. (2013). A comparison of various artificial intelligence methods in the prediction of bank failures. *Computational Economics*, 42(2) 199–215. doi: 10.1007/s10614-012-9332-0.
- Everitt, T. (2017). *What is the difference between Q-learning and R-learning?* Quora. Consultado em 2019, Março. Disponível em: <https://www.quora.com/What-is-the-difference-between-Q-learning-and-R-learning>
- Feldman, D., & Gross, S. (2005). Mortgage default: classification trees analysis. *The Journal of Real State Finance*, 30, 369-396.
- FERMA. (2018). Federation of European Risk Management Associations. *What is Risk Management?* Consultado em 2019, Janeiro. Disponível em: <https://www.ferma.eu/about/about-ferma>
- Ferreira, M. A., Celso, A. S., & Barbosa Neto, J. E. (2012). Aplicação do modelo logit binominal na análise do risco de crédito em uma instituição bancária. *Revista de Negócios*, 17(1), 41 – 59.
- FSI. (2010). Financial Stability Institute. *Regulatory use of system-wide estimations of PD, LGD and EAD*. Consultado em 2019, Janeiro. Disponível em: <https://www.bis.org/fsi/awp2010.pdf>

Flores, A., Icoy, R., Pena, C., & Gorro, K. (2018). An evaluation of SVM and Naïve Bayes with SMOTE on sentiment analysis data. In *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST) Engineering*.

Florez-Lopez, R., & Ramon-Jeronimo, J. M. (2015). Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. a correlated-adjusted decision forest proposal. *Expert Systems with Applications, Elsevier, 42(13)*, 5737–5753.

Florissi, P. (2013). *Como Big Data pode tornar empresas mais lucrativas por Déborah Oliveira*. Consultado em 2018, Dezembro. Disponível em <https://cio.com.br/o-big-data-trilha-o-caminho-da-inteligencia/>

Forti, M. (2018). *Técnicas de machine learning aplicadas na recuperação de crédito do mercado brasileiro* (Dissertação de Mestrado em Economia). Escola de Economia de São Paulo da Fundação Getúlio Vargas.

Fritz, S., & Hosemann, D. (2000). Restructuring the credit process: behaviour scoring for german corporates. *Intelligent Systems in Accounting, Finance and Management, 9*, 9-21.

Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *Computational Economics - ResearchGate*.

Garcia, M., & Giambiagi, F. (2010). *Risco e Regulação*. Rio de Janeiro: Elsevier.

Gartner Group. (2012). What is *Big data*? Gartner IT Glossary. Consultado em 2019, Janeiro. Disponível em <https://www.gartner.com/it-glossary/big-data/>

Garychl. (2018). *Applications of Reinforcement Learning in Real World*. Towards data science. Consultado em 2019, Janeiro. Disponível em <https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12>

Gaspar, C. (2014). Risco de Crédito - A importância da gestão de carteiras de crédito. *Inforbanca 100*. Instituto de Formação Bancária. Depósito Legal 15365/87, 41- 43.

Ghodselahe, A., & Amirmadhi, A. (2011). Application of artificial intelligence techniques for credit risk evaluation. *International Journal of Modeling and Optimization, 1(3)*.

Ghosh, S. & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. *Proceedings of the 27th Annual Hawaii International Conference on System Science, 3: Information Systems: DSS/Knowledge-Based Systems*, Nunamaker, J. F. & Sprague, R. H., Los Alamitos, CA, USA, 621-630.

Giryès, R., & Elad, M. (2011). *Aprendizagem por Reforço: A Survey*.

- Goldman, A., Kon, F., Pereira Junior, F., Polato, I., & Pereira, R. (2012). Apache Hadoop: Conceitos teóricos e práticos, evolução e novas possibilidades. In XXXI Jornadas de atualizações em informática, Universidade de Santa Catarina, 88-136.
- Gonçalves, E. (2005). *Análise de risco de crédito com uso de modelos regressão logística, redes neurais e algoritmos genéticos* (Dissertação de Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo.
- Gonen, G., Gonen, M., & Gargen, F. (2012). *Probabilistic and discriminative group-wise feature selection methods for credit risk analysis*. ISSN 0957-4174. Elsevier.
- Gordy, M. (2000). A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24, 119-149. doi:10.1016/S0378-4266(99)00054-0.
- Grolinger, K., L'Heureux, A., Capretz, M., & Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings*, 112, 222–233.
- Guimarães, A., & Moreira, T. (2008). Previsão de insolvência: um modelo baseado em índices contábeis com utilização da análise discriminante. *Revista de Economia Contemporânea*, 12(1), 151-178.
- Guimarães, I. A., & Chaves Neto, A. (2002). Reconhecimento de padrões: Metodologias estatísticas em crédito ao consumidor. *RAE-eletrônica*, 1(2).
- Gupton, G., Finger, C., & Bhatia, M. (1997). CreditMetrics™ — Technical Document. J.P. Morgan & Co. Incorporated.
- Härle, P., Havas, A., & Samandari, H. (2016). The future of bank risk management. *McKinsey&Company*. McKinsey&Company.
- Hair Jr, J. F., Anderson, R. E., Tatham, R., & Black, W. (2007). *Análise multivariada de dados*. Bookman.
- Han, M. (2018). Computation of loss distribution based on the structural model for credit portfolios.
- Harris, T. (2013). Quantitative credit risk assessment using support vector machine: Broad versus narrow default definitions. *Expert systems with applications*, 40, 4404-4413.
- Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42, 741–750.
- Haykin, S. (2001). *Redes Neurais: Princípios e prática*. Tradução de Paulo Martins Engel. Porto Alegre - RS: Bookman.
- He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems With Applications*, 98, 105–117.

He, J., Lan, M., Tan, C., Sung, S., & Low, H. (2004). *Initialization of cluster refinement algorithms: A review and comparative study 2004*. In Proceedings of 2004 IEEE International Joint Conference on Neural Networks.

Hens, A. B. & Tiwari, M. K. (2012). Computational time reduction for credit scoring: An integrated approach based on support vector machine and stratified sampling method. *Expert Systems with Applications*, 39(8), 6774–6781.

Hoe, D. (2013). *Como Big Data pode tornar empresas mais lucrativas por Déborah Oliveira*. Consultado em 2019, Janeiro. Disponível em <https://cio.com.br/o-big-data-trilha-o-caminho-da-inteligencia/>

Holton, G. (2004). Defining Risk. *Financial Analysts Journal*, 60.

Hsieh, N.-C. (2005). Hybrid mining approach in the design of credit scoring models. *Expert Systems with Applications*, 28, Elsevier, 655–665.

Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 33, Elsevier, ScienceDirect, 847–856.

Huang, J.-J., Tzeng, G.-H., & Ong, C.-S. (2006). Two-stage genetic programming (2SGP) for the credit scoring model. *Applied Mathematics and Computation*, 174, Elsevier, 1039–1053.

Huang, J.-Z., & Kong, W. (2003). Explaining credit spread changes: Some new evidence from option-adjusted spreads of bond indices. *Stern School of Business Working Paper No. FIN-03-013*.

Huang, K., Yang, H., King, I., & Lyu, M. (2005). Local learning vs. global learning: An introduction to maxi-min margin machine. In Support vector machines: Theory and applications. *Springer Berlin Heidelberg*, 113–131.

Huang, X., Acero, A., & Hon, H.-W. (2001). Spoken Language Processing.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, 37(4), 543–558.

Hull, K. (1992). *Risk Analysis Techniques in Defence Procurement*. Proceedings of IEEE colloquium on risk analysis methods and tools.

Ian, H. W., & Eibe Frank. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. Consultado em 2019, Janeiro. Disponível em <https://doi.org/10.1016/C2009-0-19715-5>

IDC. (2019). *IDC - Analyze the future*. A inteligência artificial ao serviço da gestão do risco de crédito. Consultado em 2019, Janeiro. Disponível em <https://www.idcdx.pt/diretorio/a-inteligencia-artificial-ao-servico-da-gestao-do-risco-de-credito/>

Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management*.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J., Ramakrishnan, R., & Shahabi, C. (2014). *Big Data and its Technical Challenges*, *Communications of the ACM*, 57(7), 86–94.

Jinsong, Y., Shuang, L., Diyin, T., & Hao, L. (2017). *Remaining Discharge Time Prognostics of Lithium-Ion Batteries Using Dirichlet Process Mixture Model and Particle Filtering Method*. IEEE.

Kadam, S., & Raval, M. (2014). Data Mining in Finance. *International Journal of Engineering Trends and Technology (IJETT)*, 16(8), ISSN:2231-5381, 377-381.

Kambal, E., Osman, I., Taha, M., Mohammed, N., & Mohammed, S. (2013). *Credit Scoring Using Data Mining Techniques with Particular Reference to Sudanese Banks*. In 2013 International Conference on Computer, Electrical and Electronics Engineering (ICCEEE).

Katsavounidis, I., Kuo, C., & Zhang, Z. (1994). *A new initialization technique for generalized Lloyd iteration*. IEEE , 144–146.

Khan, M., Durrani, M., Khalid, S., & Aziz, F. (2016). Lifelong Aspect Extraction from Big Data: Knowledge Engineering. *Complex Adaptive Systems Modeling*, 4(1), 1–15.

Kim, Y., & Sohn, S. (2012). Stock fraud detection using peer group analysis. *Expert Systems with Applications*, 39(10), 8986-8992.

Klieštík, T., & Cúg, J. (2015). Comparison of Selected Models of Credit Risk. *Procedia Economics and Finance*, 23, 356-361.

Koh, H.C., & Low, C.K. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, 19(3), 462-476. Consultado em 2019, Fevereiro. Disponível em <https://doi.org/10.1108/02686900410524436>

Kollar, B., Weissovalvana, I., & Siekelova, A. (2015). Quantification of credit risk with the use of CreditMetrics. *Procedia Economics and Finance*, 26, 311-316.

Konovalova , N., Kristovska , I., & Kudinska , M. (2016). Credit risk management in commercial banks. *Polish Journal of Management Studies*, 13(2).

Koutanaei, F., Sajedi , H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, 27, 11–23.

Kruppa, J., Schwarz, A., Armingier, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, 40, 5125-5131.

- Lagoudakis, M., & Parr, R. (2003). Least-squares policy iteration. *Journal of Machine Learning Research*, 4, 1107-1149.
- Lai, K., Yu, L., Wang, S., & Zhou, L. (2006). Credit risk analysis using a reliability based neural network ensemble model. *Artificial Neural Networks-ICANN*. Springer Berlin Heidelberg, 682–690.
- Laws, D., & A. O'Hagan. (2003). A hierarchical Bayes model for multilocation auditing. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(4), 431-450.
- Lee, T.-S., Chiu, C.-C., Lu, C.-J., & Chen, I.-F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23 - Pergamon, 245–254.
- Lee, Y.-C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, 33, Elsevier - ScienceDirect, 67–74.
- Lemos, E., Steiner, M., & Nievola, J. (2005). Análise de crédito bancário por meio de redes neurais e árvore de decisão: Uma aplicação simples de data mining. *Revista de Administração da Universidade de São Paulo*, 40(3), 225–234.
- Lempriere, M. (2017). *Telstra Enterprise*. Consultado em 2019, Janeiro. Disponível em https://www.telstraglobal.com/Creating-Transformative-Innovation-for-the-Financial-Services-Industry/assets/Transforming_financial_services_with_machine_learning_101.pdf
- Leskovec, J., Rajaraman, A., & Ullman, J. (2014). Mining of massive datasets. *Cambridge University Press*, 13.
- Leslie, P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Artificial Intelligence Research*.
- Li, J., Liu, J., Xu, W., & Shi, Y. (2004). Support vector machines approach to credit assessment. *Conference Paper in Lecture Notes in Computer Science*.
- L'Heureux, A., Grolinger, K., & Hany, F. (2017). *Machine Learning with Big Data: Challenges and Approaches*. IEEE. doi 10.1109/ACCESS.2017.2696365.
- Lisboa, I., & Miguel, J. (2018). How to manage credit risk. *European Journal of Applied Business Management*, Special Issue of ICABM2018, ISSN 2183-5594, 244-257.
- Louzada-Neto, F., & Ara, A. (2012). Modelagem de crédito via redes probabilísticas KDB. *Revista tecnológica em crédito*, 80, 20-29.
- Louzada-Neto, F., Ferreira-Silva, P., & Diniz, C. (2012). On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. *Expert Systems with Applications*, 39, 8071–8078.
- Lu, F. Q., Huang, M., Ching, W. K. & Siu, T. K. (2013) Credit portfolio management using two level particle swarm optimization. *Information Science*, 237(10), 162-175.

Lueska, L. (2009). *Modelos de Gestão de Risco de Crédito: Controle do Risco de Crédito em Instituições Financeiras* (Dissertação de Graduação em Engenharia de Produção). Escola Politécnica da Universidade de São Paulo.

Luger, G. (2004). *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos*, 4ª ed.. Porto Alegre: Bookman, 23-25. ISBN 85-363-0396-4.

Luger, G. (2009). *Artificial intelligence: structures and strategies for complex problem solving*. Boston: Pearson/Addison-Wesley.

Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, 65, 465–470.

Machado, A. (2016). *Collection scoring via regressão logística e modelo de riscos proporcionais decox* (Dissertação de Mestrado em Estatística). Instituto de Ciências Exatas da Universidade de Brasília. doi: [10.26512/2015.07.d.19575](https://doi.org/10.26512/2015.07.d.19575)

Mahadevan, S. (1996). Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine Learning*, 22.

Maher, J., & Sen, T. (1998). Predicting bond ratings using neural networks: a comparison with logistic regression. *Intelligent Systems in Accounting, Finance and Management*, 6(1), 59-72.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *The McKinsey Global Institute*.

Mardsen, P. (2017). *AI Timeline*. Digital Wellbeing. Consultado em 2019, Fevereiro. Disponível em <https://i2.wp.com/digitalwellbeing.org/wp-content/uploads/2017/08/Artificial-Intelligence-AI-Timeline-Infographic.jpeg?w=1920&ssl=1>

Markowitz, H. (1952). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, New York, 1959.

Marqués, A., García, V., & Sánchez, J. (2012). Two-level classifier ensembles for credit risk assessment. *Expert Systems with applications*, 39, 10916-10922.

Martens, D., Baesens, B., Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, 183, Elsevier - ScienceDirect, 1466–1476.

Matos, D. (2015). *Conceitos Fundamentais de Machine Learning*. Ciência e Dados. Consultado em 2019, Janeiro. Disponível em <http://www.cienciaedados.com/conceitos-fundamentais-de-machine-learning/>

- Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution that Will Transform how We Live, Work, and Think. *American Journal of Epidemiology*, 179(9).
- McAfee, A., & Brynjolfsson, E. (2012). Big Data the Management Revolution. *Harvard Business Review*. Consultado em 2019, Janeiro. Disponível em <https://hbr.org/2012/10/big-data-the-management-revolution>
- McCarthy, J. (2007). What is Artificial Intelligence? *Computer Science Department Stanford University*.
- McClelland, J. (2015). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Second Edition. Consultado em 2019, Fevereiro. Disponível em <https://web.stanford.edu/group/pdplab/pdphandbook/handbookch10.html>
- McGonagle, J., Pilling, G., & Tembo, V. (2019). *Gaussian Mixture Model*. Consultado em 2019, Fevereiro. Disponível em <https://brilliant.org/wiki/gaussian-mixture-model/>
- Melo, F. (2007). *Convergence of Q-learning: a simple proof*. Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa.
- Memon, M. A., Soomro, S., Jumani, A. K., & Kartio, M. (2017). Big Data Analytics and Its Applications. *Annals of Emerging Technologies in Computing (AETiC)*, 1(1).
- Mendes Filho, E., Carvalho, A., & Matias, A. (1996). *Utilização de Redes Neurais Artificiais na Análise de Risco de Crédito a Pessoas Físicas*. In: III Simpósio Brasileiro de Redes Neurais, Recife.
- Mendicino, C., & Punzi, M. (2013). *Confiança e Atividade Económica: O caso de Portugal*. Banco de Portugal, Boletim Económico (Inverno), 43–53.
- Michaelis. (2015). *Michaelis Dicionário Brasileiro da Língua Portuguesa*. Consultado em 2019, Janeiro. Disponível em <https://michaelis.uol.com.br/palavra/V4j7A/risco-2/>
- Mikhail, E. & Ackerman, F. (1976). *Observations and Least Squares*. University Press of America, p. 497.
- Mileris, R. (2012). Macroeconomic determinants of loan portfolio credit risk in banks. *Inzinerine Ekonomika-Engineering Economics*, 496-504.
- Minussi, J. A., Damacena, C., & Ness Jr., W. L. (2002). Um modelo de previsão de solvência utilizando regressão logística. *Revista de Administração Contemporânea*, 6(3), 109-128.
- Misankova, M., Kocisova, K., Frajtova-Michalikova, K., & Adamko, P. (2014). CreditMetrics and its use for the calculation of credit risk. In 2nd International Conference on Economics and Social Science. *Information Engineering Research Institute, Advances in Education Research*, 61, 124-129.

- Mishra , M., & Dash , R. (2014). *A Comparative Study of Chebyshev Functional Link Artificial Neural Network, Multi-Layer Perceptron and Decision Tree for Credit Card Fraud Detection*. In 13th International Conference on Information Technology. doi: 10.1109/icit.2014.25
- Mitchell, T. (1997). *Machine Learning*. p. 2. ISBN 978-0-07-042807-2. McGraw Hill.
- Monico, J. F., Póz, A. P., Galo, M., Santos, M. C., & Oliveira, L. C. (2009). Acurácia e Precisão: Revendo os Conceitos de Forma Acurada. *Boletim de Ciencias Geodesicas*.
- Morais, J. (2011). *Risco de Crédito e o Efeito dos Colaterais*. Faculdade de Economia da Universidade de Coimbra.
- Mun, J. (2004). *Applied Risk Analysis*. John Wiley & Sons.
- Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J., & Ryu, K. (2019). An empirical comparison of machine learning methods on bank client credit Assessments. *Sustainability Journal*.
- Na, S., Xumin, L., & Yong , G. (2010). *Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm*.
- Naik, U., & Shivalingaiah, D. (2008). *Comparative Study of Web 1.0, Web 2.0 and Web 3.0*. In 6th International Caliber, Allahabad, 499-507.
- Nanda, S., & Pendharkar, P. (2001). Linear models for minimizing misclassification costs in bankruptcy prediction. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 3(10), 155-168.
- Nascimento, D. (2017). *Machine Learning - entendendo o processo de aprendizagem*. Medium. Consultado em 2019, Fevereiro. Disponível em https://medium.com/@Daniel_nasci/machine-learning-entendendo-o-processo-de-aprendizagem-80835b3ec2dc
- Neves, M. E., & Quelhas, A. P. (2013). *Carteiras de Investimento - Gestão e Avaliação do Desempenho*. Coimbra: Almedina.
- Ngai, E., Y. Hu, Y. Wong, Y. Chen, & X. Sun. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569.
- Nikolic, N., Zarkic-Joksimovic, N., Stojanoviski, D., & Joksimovic, I. (2013). The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems With Applications*, 40, 5932-5944.
- Nordstrom, E., & Carlstrom, J. (1997). *Near-optimal link allocation of blockable narrow-band and queueable wide-band call traffic in ATM networks*. In V. Ramasawami and P. E. Wirth, editors, Proceedings of the 15th International Teletraffic Congress, ITC'15 - Elsevier Science.

- Olson, D., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 52, 464-473.
- Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications*, 29, Elsevier - ScienceDirect, 41-47.
- Oxford University Press. (2005). Risco. *Dictionary of Finance and Banking*. 3a Edição. Oxford: Oxford University Press.
- Pardo, B., & Birmingham, W. (2005). *Modeling Form for On-line Following of Musical Performances*. AAAI.
- Pascual, M. B., Martínez, A. M., & Alamillos, A. M. (2014). Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica. *Cuadernos de Economía*, 37(104), 73-86.
- Perera, L. (2013). Crédito. *História, fundamentos e modelos de análises*.
- Picinini, R., Oliveira, G., & Monteiro, L. (2003). *Mineração de critério de credit scoring utilizando algoritmos genéticos*. Simpósio Brasileiro de Automação Inteligente, 6.
- Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research*, 156, Elsevier - ScienceDirect, 483-494.
- Pires, I. (2008). *Um modelo estratégico para análise de crédito utilizando redes neurais artificiais* (Dissertação de Mestrado). Universidade Federal de Uberlândia.
- Poudel, R. (2013). *Macroeconomic Determinants of Credit Risk in Nepalese Banking Industry*. In Proceedings of 21st International Business Research Conference 10 - 11 June, 2013, Ryerson University, Toronto, Canada, ISBN: 978-1-922069-25-2.
- Protil, R., & Silva, W. (2006). *Utilização da função de análise discriminante linear e o modelo de regressão logística na previsão de insolvência de cooperativas agrícolas do estado do Paraná*. In XLIV CONGRESSO DA SOBER - Sociedade Brasileira de Economia e Sociologia Rural, Fortaleza.
- Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). *A survey of machine learning for big data processing*. In *EURASIP*.
- Rajamohamed, R., & Manokaran, J. (2018). Improved credit card churn prediction based on rough clustering and supervised learning techniques. *Cluster Computing*, 21(3) 65-77. doi: 10.1007/s10586-017-0933-1
- Ray, S. (2017). *6 Easy Steps to Learn Naive Bayes Algorithm (with codes in Python and R)*. Consultado em 2019, Março. Disponível em <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>

- Ribeiro, C. F., Zani, J., & Zanini, F. A. (2009). *Estimação da probabilidade de inadimplência: uma verificação empírica na universidade católica de pelotas*. Consultado em 2019, Janeiro. Disponível em <http://www.anpad.org.br/admin/pdf/FIN645.pdf>
- Rich, E., Knight, K., & Nair, S. (2009). *Artificial Intelligence*. Tata McGraw-Hill. Consultado em 2019, Janeiro. Disponível em <https://i4iam.files.wordpress.com/2013/08/artificial-intelligence-by-rich-and-knight.pdf>
- Rosa, P. (2000). *Modelos de Credit Scoring Regressão Logística Chaid e Real* (Dissertação de Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo.
- Rouse, M. (2018). *Machine Learning Definition*. Consultado em 2019, Janeiro. Consultado em <http://whatis.techtarget.com/definition/machine-learning>
- Russell, S., & Norvig, P. (2013). *Inteligência Artificial*. Elsevier - Campus.
- Sabbeh, S. (2018). *Machine Learning Techniques for Customer Retention: A Comparative Study*. In (IJACSA) International Journal of Advanced Computer Science and Applications, 9(2).
- Sabzevari, H., Soleymani, M., & Noorbakhsh, E. (2007). *A comparison between statistical and data mining methods for credit scoring in case of limited available data*.
- Salgado, S. (2015). *Estudo Comparativo entre Regressão Logística Binária e Redes Neurais Artificiais na Avaliação dos Resultados Clássicos de Hosmer, Lemeshow e Sturdivant* (Dissertação de Mestrado), Universidade Federal de Itajubá.
- Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*.
- Santhosh, K., Dogra, D., & Roy, P. (2017). *Real-Time moving object classification using DPMM for road traffic management in smart cities*. IEEE.
- Santos, J. (2011). *Metodologia de implementação do método das notas internas* (Dissertação de Mestrado em Engenharia Industrial e Gestão). Faculdade de Engenharia da Universidade do Porto.
- Santosh, D., Venkatesh, P., Poornesh, P., Rao, L., & Kumar, N. (2013). Tracking multiple moving objects using gaussian mixture model. *International Journal of Soft Computing and Engineering (IJSCE)*.
- Sarkar, S., & Sriram, R. (2001). Bayesian models for early warning of bank failures. *Management Science*, 47(11), 1457-1475.
- Satish L, G. (2003). *Use of hidden Markov models for partial discharge pattern classification*. IEEE.

- Saunders, A. (1999). Credit risk measurement: value-at-risk and other new paradigms. *John Wiley & Sons*.
- Saunders, A., & Cornett, M. (2011). *Financial Markets and Institutions - A Risk Management Approach*. 7th Edition, McGraw Hill Irwin, New York.
- Schade, G. (2018). *IA, ML, DL e RNA não são a mesma coisa!* Consultado em 2019, Fevereiro. Disponível em <https://gabrielschade.github.io/2018/03/26/IA-concepts.html>
- Schwartz, A. (1993). *A reinforcement learning method for maximizing undiscounted rewards*. In proceedings of the tenth international conference on machine learning, 298-305.
- Segal, T. (2019). *Spread Definition: Whats is a spread?* Consultado em 2019, Fevereiro. Disponível em <https://www.investopedia.com/terms/s/spread.asp>
- Selau, L. P. (2008). *Construção de Modelos de Previsão de Risco de Crédito* (Dissertação de Mestrado em Engenharia de Produção). Universidade Federal do Rio Grande do Sul.
- Semolini, R. (2002). *Support Vector Machines, Inferência Transdutiva e o Problema de Classificação* (Tese de Mestrado). Faculdade de Engenharia Elétrica e de Computação, Unicamp.
- Silva, E. S., Mota, C., Queirós, M., & Pereira, A. (2013). Finanças e Gestão de Riscos Internacionais. *Vida Económica*, 228. Porto.
- Silva, R. A. (2014). *Performance financeira da carteira na avaliação de modelos de análise e concessão de crédito: uma abordagem baseada em aprendizagem estatística* (Tese de Doutorado em Ciências). Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto. Universidade de São Paulo.
- Silva, R., Ribeiro, E., & Matias, A. (2016). Aprendizagem estatística aplicada à previsão de default de crédito. *Revista de Finanças Aplicadas*.
- Silver, D., Yang, Q., & Li, L. (2013). Lifelong Machine Learning Systems: Beyond Learning Algorithms. *AAAI Spring Symposium Series*, 49–55.
- Singh, S., & Jaiswal, U. (2018). Machine Learning for Big Data: A New Perspective. *International Journal of Applied Engineering Research*, 13(5), 2753-2762. ISSN 0973-4562
- Smolyakov, V. (2017). *Bayesian Nonparametrics. An introduction to the Dirichlet process and its applications*. Medium. Consultado em 2019, Fevereiro. Disponível em <https://blog.statsbot.co/bayesian-nonparametrics-9f2ce7074b97>
- Sousa, M., & Figueiredo, R. (2014). Análise de crédito por meio de mineração de dados: Aplicação em cooperativa de crédito. *JISTEM - Journal of Information Systems and Technology Management, Revista de Gestão da Tecnologia e Sistemas de Informação* 11(2), 379-396.

Souza, R. (2000). *O modelo de collection scoring como ferramenta para a gestão estratégica do risco de crédito* (Dissertação de Mestrado em Administração). Fundação Getúlio Vargas. Escola de Administração de Empresas de São Paulo.

Spathis, C. (2002). Detecting false financial statements using published data: Some evidence from Greece. *Managerial Auditing Journal*, 17(4), 179-191.
doi: 10.1108/02686900210424321

Sriram, R. (1997). *Intelligent Systems for Engineering - A Knowledge-based Approach*. Springer Science & Business Média. doi:10.10007/978-1-4471-0631-9

Steiner, M., Nievola, J., Soma, N., Shimizu, T., & Steiner Neto, P. (2007). Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesquisa Operacional*, 27(3), 407-426.

Stratonovich, R. (1960). Conditional Markov Processes. *Theory of Probability and its Applications*, 5(2), 156–178. doi:10.1137/1105015

Sutton, R. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1), 9–44. doi:10.1007/BF00115009.

Sutton, R., & Barto, A. (2017). *Reinforcement Learning: An Introduction Second edition*. A Bradford Book. The MIT Press. Cambridge, Massachusetts. London, England.

Swiderski, B., Kurek, J., & Osowski, S. (2012). Multistage classification by using logistic regression and neural networks for assessment of financial condition of company. *Decision Support Systems*, 52(2), 539–547.

Syeda, M., Zhang, Y-Q. & Pan, Y. (2002). Parallel granular neural networks for fast credit card fraud detection. In *Proceedings of the 2002 IEEE International Conference*, 1, 572-577.

Tadepalli, P., & Ok, D. (1994). *H-Learning: A reinforcement learning method to optimize undiscounted average reward*. Department of Computer Science, Oregon State University.

Taleb, N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Ediciones Paidós Ibérica S.A. ISBN: 9788449321894

Tam, K. Y.; Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management Science, Inform*s, 38(7), 926–947.

Tambet, M. (2015). *Demystifying Deep Reinforcement Learning*. Computational Neuroscience Lab. Institute of Computer Science. University of Tartu. Consultado em 2019, Março. Disponível em <https://neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/>

Tarihi, G., & Tarihi, K. (2016). Comparison of machine learning techniques for analyzing banks financial distress. *The Journal of Social Sciences Institute*, 19(36). Balıkesir University.

Techopedia. (2019). *Definition - What does Gaussian Mixture Model (GMM) mean?* Techopedia. Consultado em 2019, Março. Disponível em <https://www.techopedia.com/definition/30331/gaussian-mixture-model-gmm>

Tecnologia e Segurança. (2012). *O risco do Big Data é a miopia*. Consultado em 2018, Dezembro. Disponível em <https://tecnologiaseguracao.wordpress.com/2012/02/27/o-risco-do-big-data-e-a-miopia/>

Thomas, L., Crook, J., & Edelman, D. (2017). *Credit Scoring and Its Applications*, Second Edition. SIAM News. ISBN 978-1-611974-55-3

Thota, S. (2017). *The Evolution of Machine Learning*. Synectics. Consultado em 2019, Fevereiro. Disponível em <https://smdi.com/2017/12/the-evolution-of-machine-learning/>

Torrey, L., & Shavlik, J. (2010). *Transfer Learning*. Handbook of Research on Machine Learning Applications and Trends. IGI Global. University of Wisconsin, Madison WI, USA. doi:10.1.1.146.1515

Trevino, A. (2016). *Introduction to K-means Clustering*. Learn Data Science, Machine Learning. Consultado em 2019, Fevereiro. Disponível em <https://www.datascience.com/blog/k-means-clustering>

Tsai, C.-F. (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems, 22*, Elsevier - ScienceDirect, 120–127.

Tsai, C.-F., & Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing, 10*, 374–380.

Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 34*, Elsevier - ScienceDirect, 2639–2649.

Tsitsiklis, J., & Van Roy, B. (1996). An Analysis of Temporal-Difference Learning with Function Approximation. *IEEE Transactions on automatic control, 42*(5), 674-690.

Turkson, R., Baagyere, E., & Wenya, G. (2016). *A Machine Learning Approach for Predicting Bank Credit Worthiness*. Consultado em 2019, Fevereiro. Disponível em <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7585216&tag=1>

US Patent. (2015). *Methods and Apparatus for Reinforcement Learning*, US Patent #20150100530A1. US Patent Office. Consultado em 2019, Fevereiro. Disponível em <https://patents.google.com/patent/US20150100530A1/em>

Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance, 22*, Elsevier - ScienceDirect, 1421-1439.

Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications, 38*, 223–230.

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems, 26*, 61–68.

Wang, Y., Wang, S., & Lai, K. (2005). A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*, 13(6).

Washington, G. (2014). Effects of Macroeconomic Variables on Credit Risk in the Kenyan Banking System. *Asian Society of Business and Commerce Research. International Journal of Business and Commerce*, 3(9), 2225-2436.

Watkins, C. (1989). *Learning from delayed rewards* (Dissertação de Doutorado). University of Cambridge.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27, 1131-1152.

Witten, I., Frank, E., Hall, M., & Pal, C. (2017). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann is an imprint of Elsevier.

Witzany, J. (2017). *Credit risk management: Pricing, measurement, and modeling*. Springer International Publishing AG. ISBN 978-3-319-49800-3

Yu, H., Huang, X., Hu, X., & Cai, H. (2010). *A Comparative Study on Data Mining Algorithms for Individual Credit Risk Evaluation*. In 2010 International Conference on Management of e-Commerce and e-Government.

Yu, L., Wang, S., & Lai, K. (2009). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. *European Journal of Operational Research*, 195, 942–959.

Yu, L., Yue, W., Wang, S., & Lai, K. (2010). Support vector machine based multiagent ensemble learning for credit risk evaluation. *Expert Systems with Applications*, 37, 1351–1360.

Yufeng, K., Chang-Tien, L., Sirwongwattana, S., & Yo-Ping, H. (2004). Survey of fraud detection techniques. *IEEE International Conference on Networking, Sensing and Control*, 2, 749-754.

Zamore, S., Djan, K., Alon, I., & Hobdari, B. (2018). Credit Risk Research: Review and Agenda. *Emerging Markets Finance and Trade*, 54(4). doi: 10.1080/1540496X.2018.1433658, 811-835.

Zheng, J., Shen, F., Fan, H., & Zhao, J. (2012). An online incremental learning support vector machine for large-scale data. *Neural Computing and Applications*, 22(5), 1023-1035.

Zikopoulos, P., Lightstone, S., Huras, M., & Sachedina, A. (2013). New dynamic in-memory analytics for the era of Big Data. *IBM Data Management Magazine*, 1–47.

Anexo

Anexo A – Questionário aos Técnicos de Crédito

ISCTE IUL - UNIVERSITY INSTITUTE OF LISBON - Research Questionnaire

Dear,

I'm master's student André Luiz Monteiro, and this questionnaire is part of a research of completion of a course at ISCTE.

This will address the value that Machine Learning can provide the institution banking in Credit Risk.

I guarantee the anonymity and confidentiality of the answers.

And thank you in advance for your cooperation!

1. Age range :

- Up to 25 years
- From 25 to 34 years
- From 35 to 44 years
- From 45 to 54 years
- Over 60 years

2. Last course that you have completed :

- PhD
- Master's degree
- Specialization
- 2nd cycle or undergraduate Level
- 1st cycle or Medium Level
- Other (Please specify)

3. Your position in company :

- Direction
 - Management
 - Coordination or leadership
 - Analyst
 - Technical
-

4. Time when you're in the company:

- 1 year or less
- more than 1 to 3 years
- more than 3 to 5 anos
- more than 5 to 10 anos
- more than 10 years

5. Number of employees in your company:

- up to 20
- more than 20 to 100
- more than 100 to 500
- more than 500 to 1000
- more than 1000

6. Your company's turnover (Millions of €):

- up to 2
- more than 2 to 10
- more than 10 to 50
- more than 50

7. What is the percentage of businesses in credit in relation to the total of the company.

- up to 20%
- more than 20 to 50 %
- more than 50 to 80 %
- more than 80 %

8. Do you participate in any activity related to Machine Learning, Big Data or IT Infrastructure?

- Yes
- No

9. Machine Learning will bring significant benefits to your company's business.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

10. Machine Learning is a technology that will be present in almost all new software products and services purchased.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

11. You consider that your company operates in Machine Learning.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

12. I believe my company is thinking of investing in at least one Machine Learning project by 2020.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

13. Shortage of Qualified Professionals in Machine Learning is one of the problems of your company.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

14. Most of the executives at your company understand the potential benefits of data-based decision making and Machine Learning technology.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

15. In understanding the benefits of Machine Learning, the executives of your company believe that this change of mindset is a long and slow process.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

16. Your company feels they need to improve data-level processes (quality and data integration) before they can fully utilize Machine Learning.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

17. Outsourcing IT infrastructure services to Machine Learning solutions is one of your company's strategies.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

18. Order your company preference for development in Machine Learning projects:

Instruções de pergunta: Change the order according to your preference (1 - most important, last - least important)

Fraud Detection	<input type="checkbox"/>
Credit Risk	<input type="checkbox"/>
Identification of Security Attacks	<input type="checkbox"/>
Predictive Maintenance	<input type="checkbox"/>
IT Systems Anomaly Detection	<input type="checkbox"/>
Forecast of Demand	<input type="checkbox"/>
Financial Negotiation	<input type="checkbox"/>

19. The credit risk of your company already uses of Machine Learning or thinking about using it.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

20. In what areas of credit is used or intends to use Machine Learning.

Instruções de pergunta: Select one or more answers

- Approval or Rejection
- Credit Monitoring
- Calculation of capital
- Other ...

21. What are the biggest challenges to implementing Machine Learning in Credit Risk in your company.

Instruções de pergunta: Select one or more answers

- Data
- Technology
- Persons
- Process
- Other...

22. Machine Learning algorithms can assist analysts in credit analysis and lending, giving them security and support during decision making.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

23. We are currently faced with increasingly high default rates. With the implementation of Machine Learning in the credit analysis sector, is it possible to affirm that there will be improvement in the rates of compliance of the customers of the company?

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

24. Due to the lack of real guarantees, banks need to obtain more information from customers that can be used as risk reduction subsidies. In this way, he believes that the use of Machine Learning can help in the qualitative analysis of the credit C's (debtor character, legal capacity, condition of the environment, collateral of the assets and financial control of the debtor).

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

25. The liquidity, solvency, and profitability ratios are indicators that can be improved and controlled by Machine Learning techniques.

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

26. In the context of Credit Risk, sort out which Machine Learning algorithms have brought / brought more effectiveness to your business:

Instruções de pergunta: *Change the order according to your preference (1- most important, last - least important)*

Decision trees	<input type="text"/>
Naïve Bayes Classification	<input type="text"/>
Linear Regression of Least Squares	<input type="text"/>
Logistic regression	<input type="text"/>
Support Vector Machine - SVM	<input type="text"/>
Ensemble Methods	<input type="text"/>
Clustering Algorithms	<input type="text"/>
Singular Values in Decomposition - SVD	<input type="text"/>
Principal Component Analysis - PCA	<input type="text"/>
Independent Components Analysis - ICA	<input type="text"/>
Artificial Neural Networks - ANN	<input type="text"/>

27. Final comments

Instruções de pergunta: *Please write below additional comments that you would like to make about this questionnaire.*

Anexo B – Questionário aos Executivos e Diretores de Crédito

Validation Questionnaire

Your Excellency, the Director(a) of Credit,

I am conducting an investigation, within the framework of a Masters degree approved by the University Institute of Lisbon, whose theme is:

Machine learning in the process of credit risk of banking institutions.

One of the main objectives of this study is about the current use of the banking industry, in machine learning for Credit Risk.

In order to achieve the above mentioned objective, I request the validation of the results obtained from the research done to the bank credit analysts, through the answer of this questionnaire.

Thank you in advance for your understanding and availability

Best regards,

André Luiz Monteiro (almoz@iscte.pt)

1. Does your company have more than 1000 employees?

- Yes
 No

2. Volume of business is over 50 Million €?

- Yes
 No

3. What is the percentage of businesses in credit about the total of the company?

- Up to 20%
 From 20 to 50%
 From 50 to 80%
 Over 80%

4. Does your company's credit risk industry already use Machine Learning?

- Yes
 No
 If not, please explain:

5. In the survey made in early 2019 to some banking credit analysts, it was diagnosed that more than 70% of respondents chose credit risk as the area of the company to develop machine learning projects. Do you agree with this statement?

- Yes
 No
 If not, please explain:

6. In this same survey to bank credit analysts, it was found that the credit approval process is preferred for use of machine learning. Is this statement correct for your business?

- Yes
 No
 If not, what would be your company's favorite credit process?

7. If you had to choose one of Data, Technology, People or Processes, what would be the biggest challenge to implement Machine Learning in Credit Risk in your company? Justify.

8. In the survey of bank credit analysts, 85% of them answered that machine learning algorithms can assist them in lending activities, giving them more security, support and agility in decision-making. Does your company agree with this thinking?

- Yes
 No
 Justify:

9. Most bank credit analysts have responded that it is possible for default rates to decline with the deployment of machine learning in the credit analysis industry. Does your company share this idea?

Yes

No

Justify:

10. Due to the lack of real guarantees, banks need to obtain more information from customers that can be used as risk reduction subsidies. In this way, he believes that the use of Machine Learning can help in the qualitative analysis of the credit C's (debtor character, legal capacity, condition of the environment, collateral of the assets and financial control of the debtor).

Instruções de pergunta: 1 - Disagree 2 - Partially Disagree 3 - Indifferent 4 - Partially Agree 5 - Agree

☆☆☆☆☆ / 5

11. In the context of the bank credit analysts research, Decision Tree, Naive Bayes and Regression (both linear and logistic) algorithms are the most effective for banks' credit business. Does your company agree with this statement?

Yes

No

If not, what would be the most effective algorithms for your business?

12. Final comments.

Instruções de pergunta: Please write below additional comments that you would like to make about this questionnaire.

Anexo C – Outras Tabelas e Figuras

Tabela 17 – Estatística de preenchimento do questionário

Banco	2 a 5 minutos	5 a 10 minutos	10 a 30 minutos	>30 minutos	% incompletos	% do total dos respondidos	Total de respostas
1	36%	36%	27%	-	21%	15,0%	11
2	44%	44%	-	11%	31%	12,3%	9
3	25%	50%	13%	13%	20%	10,9%	8
4	36%	54%	10%	-	27%	15,0%	11
5	-	83%	17%	-	40%	8,2%	6
6	-	67%	33%	-	71%	9,5%	7
7	33%	50%	17%	-	33%	8,2%	6
8	100%	-	-	-	67%	5,4%	4
9	-	75%	25%	-	33%	5,4%	4
10	33%	67%	-	-	-	4,1%	3
11	100%	-	-	-	60%	2,7%	2
12	50%	50%	-	-	33%	2,7%	2

Fonte: Elaboração própria.

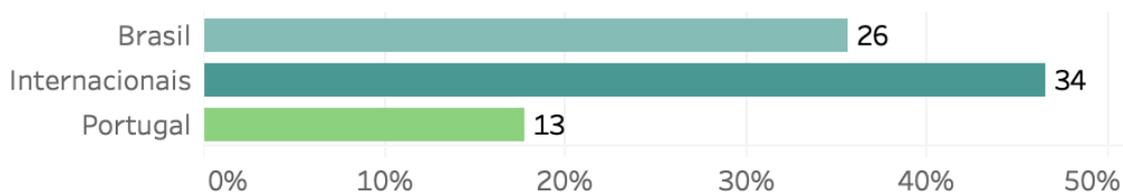


Figura 50 – Percentual de respostas por nicho.

Fonte: Elaboração própria.

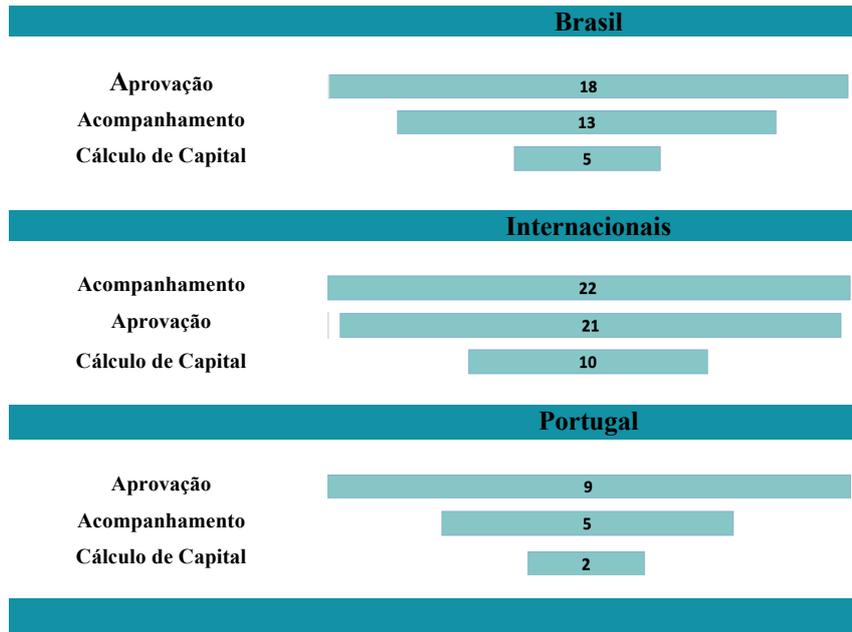


Figura 51 – Quais áreas do crédito para utilização de ML por nichos?

Fonte: Elaboração própria.

	Nichos		
	Brasil	Internacionais	Portugal
Pessoas	16 pts	10 pts	8 pts
Tecnologia	11 pts	17 pts	6 pts
Processos	11 pts	16 pts	6 pts
Dados	13 pts	13 pts	4 pts

Figura 52 – Desafios de ML em RC por nichos.

Fonte: Elaboração própria.

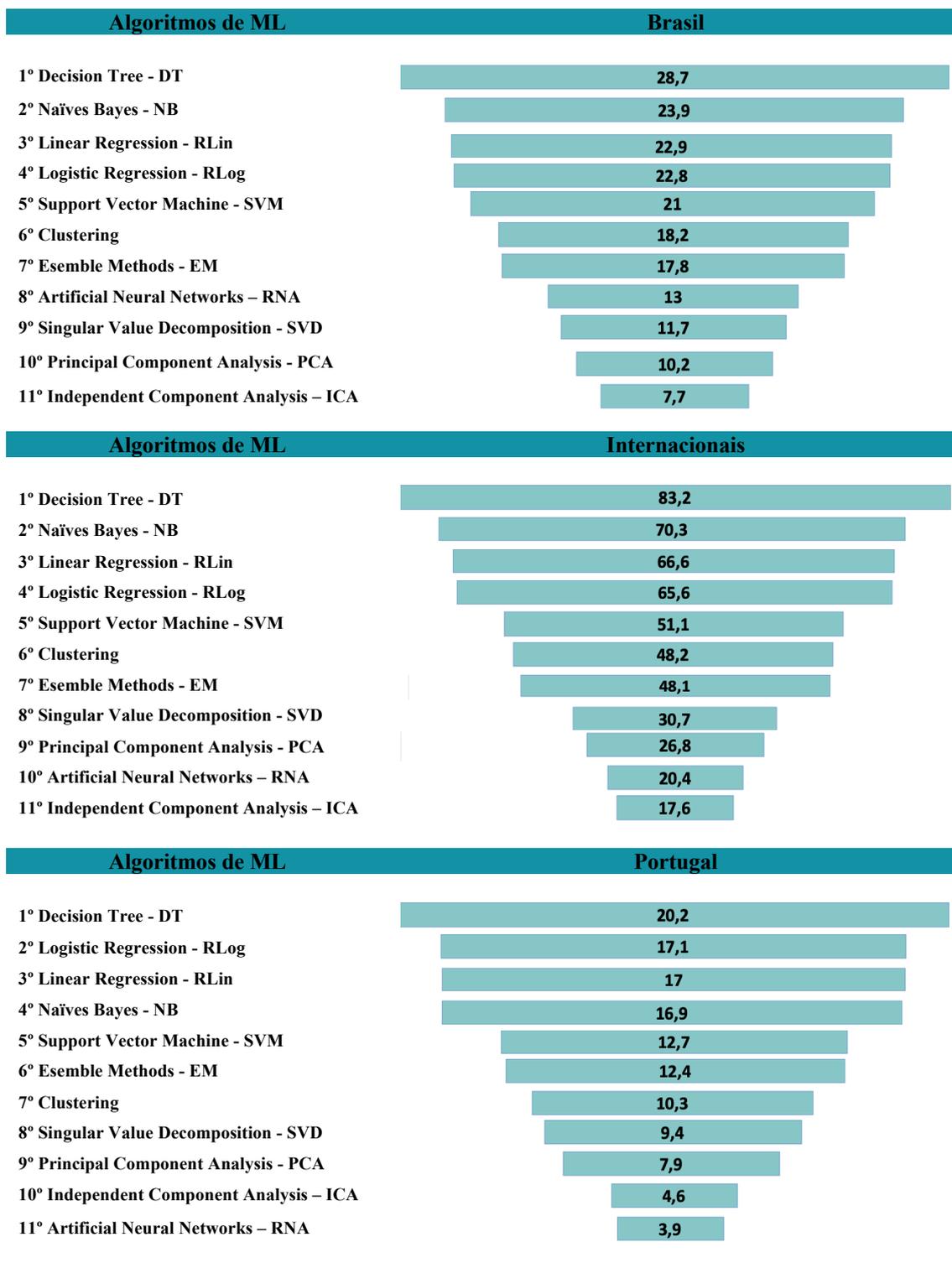


Figura 53 – Os algoritmos de ML e eficácia para os negócios por nichos.

Fonte: Elaboração própria.

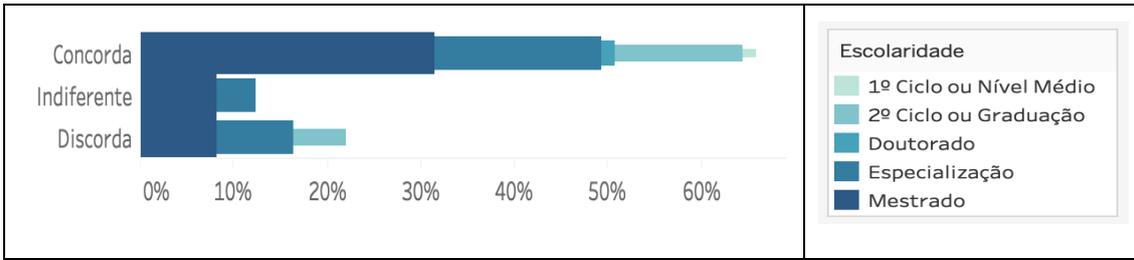


Figura 54 – A perspectiva de uso de ML em RC por escolaridade.

Fonte: Elaboração própria.

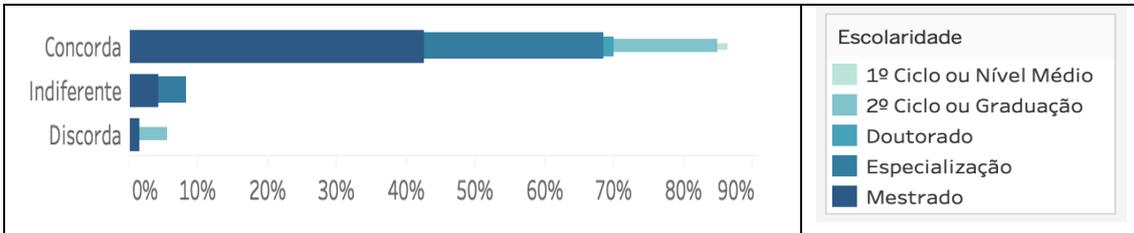


Figura 55 – ML podem auxiliar análises e concessões de crédito por escolaridade?

Fonte: Elaboração própria

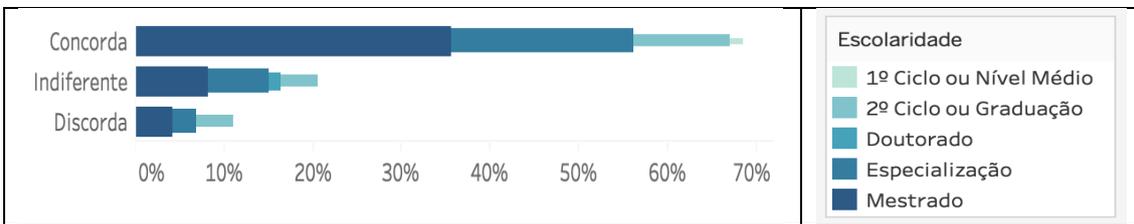


Figura 56 – ML e os índices de incumprimento por escolaridade.

Fonte: Elaboração própria.

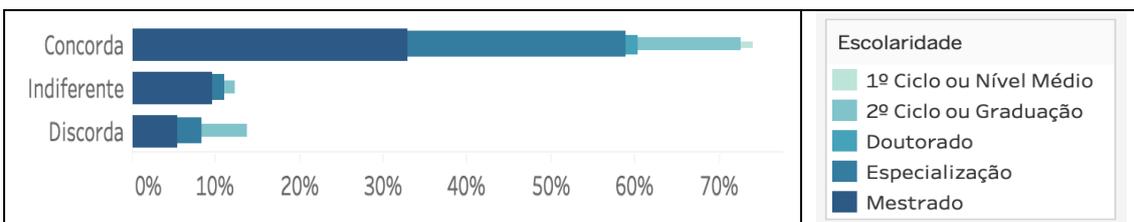


Figura 57 – ML na análise qualitativa do crédito por escolaridade.

Fonte: Elaboração própria.

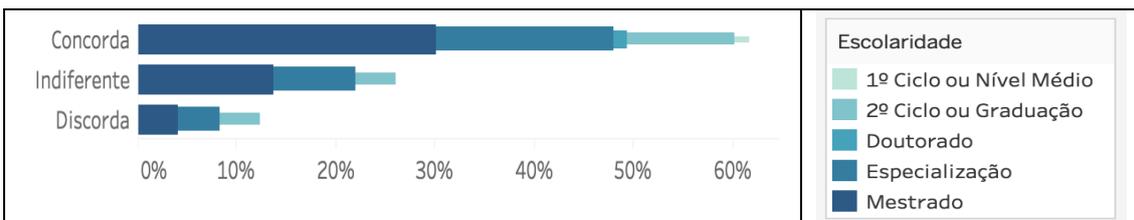


Figura 58 – ML nos RL, solvabilidade e rendibilidade por escolaridade.

Fonte: Elaboração própria.

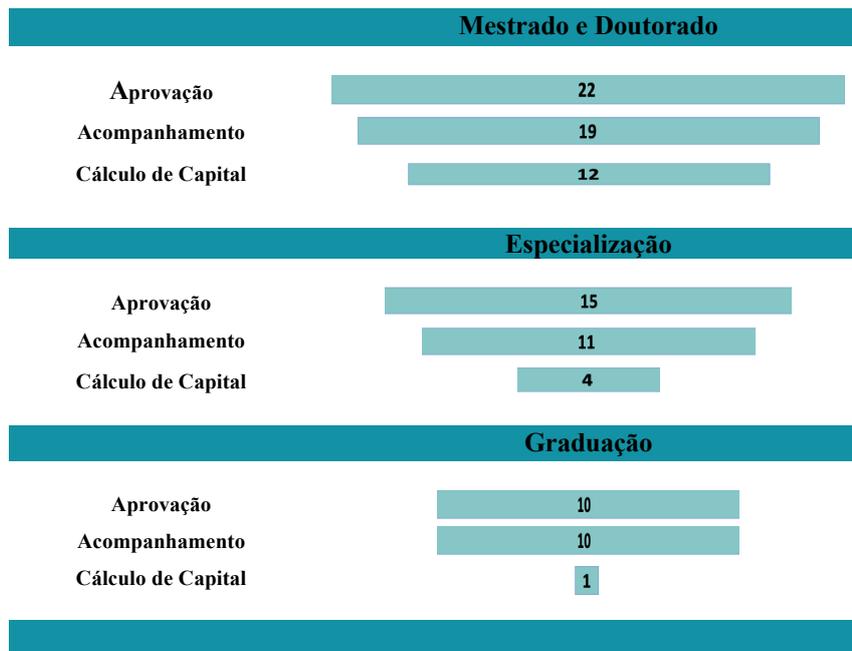


Figura 59 – Quais áreas do crédito para utilização de ML por escolaridade?

Fonte: Elaboração própria.

	Escolaridade				
	1º Ciclo ou Nível Médio	2º Ciclo ou Graduação	Especialização	Mestrado	Doutorado
Pessoas	1 pts	6 pts	12 pts	14 pts	1 pts
Tecnologia	0 pts	5 pts	11 pts	17 pts	1 pts
Processos	1 pts	8 pts	6 pts	18 pts	0 pts
Dados	0 pts	6 pts	10 pts	13 pts	1 pts

Figura 60 – Desafios de ML em RC por escolaridade.

Fonte: Elaboração própria.

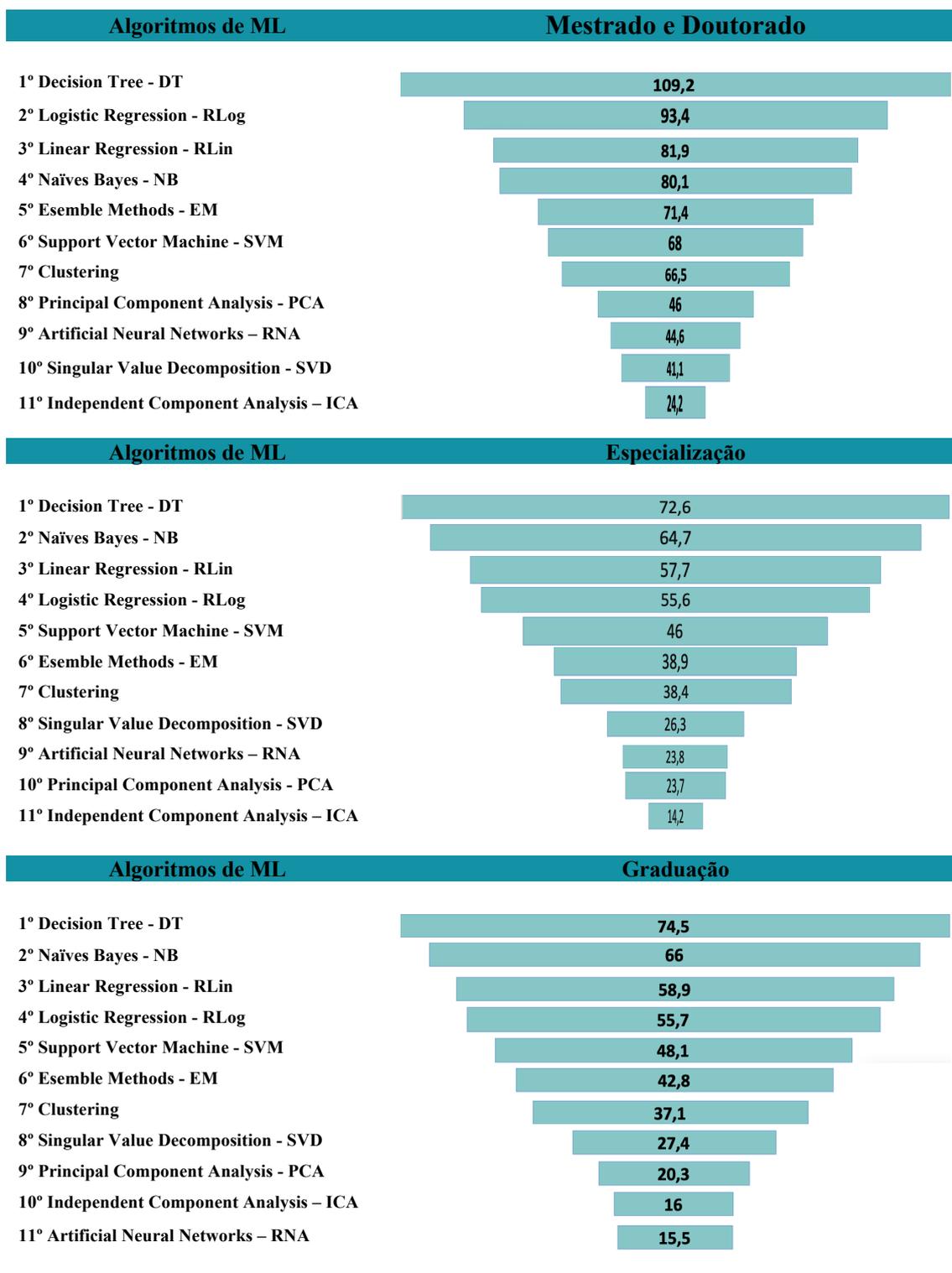


Figura 61 – Os algoritmos de ML e eficácia para os negócios por escolaridade.

Fonte: Elaboração própria.