



Universidade de Brasília

Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas

Departamento de Administração

YASMIN MARCELLE BARBOSA FREIRE

**PONTUAÇÃO DE CRÉDITO COM USO DE BIG DATA: 10
ANOS DE REVISÃO SISTEMÁTICA MEDIANTE ENFOQUE
META ANALÍTICO**

Brasília – DF

2019

YASMIN MARCELLE BARBOSA FREIRE

**PONTUAÇÃO DE CRÉDITO COM USO DE BIG DATA: 10
ANOS DE REVISÃO SISTEMÁTICA MEDIANTE ENFOQUE
META ANALÍTICO**

Monografia apresentada ao
Departamento de Administração como
requisito parcial à obtenção do título de
Bacharel em Administração.

Professor Orientador: Carlos Henrique
Rocha, PhD

Brasília – DF

2019

Freire, Yasmin Marcelle Barbosa.

Pontuação de Crédito com uso de *Big Data*: 10 anos de revisão sistemática mediante enfoque meta analítico / Yasmin Marcelle Barbosa Freire – Brasília, 2019.

46 f. : il.

Monografia (bacharelado) – Universidade de Brasília, Departamento de Administração, 2019.

Orientador: Prof. Carlos Henrique Rocha, Departamento de Administração.

1. Pontuação de Crédito. 2. Big Data. 3. Enfoque Meta Analítico. 4. Bibliometria I. Título.

YASMIN MARCELLE BARBOSA FRIERE

**PONTUAÇÃO DE CRÉDITO COM USO DE BIG DATA: 10
ANOS DDE REVISÃO SISTEMÁTICA MEDIANTE ENFOQUE
META ANALÍTICO**

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de
Conclusão do Curso de Administração da Universidade de Brasília do
(a) aluno (a)

Yasmin Marcelle Barbosa Freire

PhD em Economia, Carlos Henrique Rocha
Professor-Orientador

Prof^a. Dr^a. Ducineli Régis Botelho
Professor-Examinador

MSc. Olívia Fernanda Rocha de
Oliveira
Professor-Examinador

Brasília, 16 de novembro de 2019

AGRADECIMENTOS

Agradeço a meu professor orientador que me deu as diretrizes para aprovação.

Dedico a meus pais que estavam ao meu lado sempre.

RESUMO

O objetivo deste estudo foi identificar a evolução de pesquisas sobre pontuação de crédito usando Big Data, analisando lacunas e limitações nos últimos dez anos. A metodologia adotada foi a revisão sistemática da literatura, de caráter exploratório por meio do enfoque meta analítico. Foi realizada a busca com os termos “*credit scoring*” (pontuação de crédito) e “*Big Data*”, no espaço temporal de dez anos (2009 - 2018), e dentro das quatro categorias que mais publicaram artigos relacionados ao assunto estudado. A busca no Web of Science resultou em vinte e quatro textos, entre artigos, *reviews* e *proceeding papers*. Os resultados sugerem que a pontuação de crédito com uso de *Big Data* é um tema novo, mas apresenta interesse acadêmico crescente em comparação a outros temas na área de finanças. A análise de autores e artigos permitiu visualizar que a produção de artigos é dispersa entre países, a produtividade tem sido crescente quando se trata do assunto em questão, e os autores com maiores citações publicaram somente o artigo mais citado sobre o assunto, sendo a maioria dos trabalhos escritos por mais de um autor. Por fim, as linhas de pesquisa estão bastante voltadas para a melhoria de decisões de empresas e instituições financeiras, inclusão social e melhoria no desempenho dos modelos de pontuação de crédito.

Palavras-chave: Pontuação de Crédito, *Big Data*, enfoque meta analítico, bibliometria

ABSTRACT

The objective of this study is to evaluate the evolution of research in the field of Big Data credit scoring in the last ten years. The methodology adopted was the systematic literature review, exploratory character through the meta-analytical approach. The search was carried out with the terms “credit scoring” and “Big Data”, within ten years (2009 - 2018), and within the four categories that most published articles related to the subject studied. The Web of Science search resulted in twenty-four texts, including articles, reviews, and proceeding papers. The results suggest that the credit score using Big Data is a new subject, but it has a growing academic interest compared to other finance subjects. The analysis of authors and articles showed that the production of articles is dispersed among countries, productivity has been increasing when it comes to the subject in question, and the authors with higher citations published only the most cited article on the subject, most of them of works written by more than one author. Finally, the lines of research are largely focused on improving business decisions and financial insights, social inclusion, and improving the performance of credit score models.

Keyword: Big Data, Focused analytical goal, Bibliometrics, Credit scoring

SUMÁRIO

1. INTRODUÇÃO	11
1.1. CONTEXTUALIZAÇÃO	11
1.2. PROBLEMA DE PESQUISA	12
1.3. OBJETIVOS	12
1.3.1. Objetivo Geral	12
1.3.2. Objetivos Específicos.....	13
1.4. JUSTIFICATIVA	13
2. REFERENCIAL TEÓRICO	14
2.1. RISCO DE CRÉDITO	14
2.2. PONTUAÇÃO DE CRÉDITO	15
2.2.1. Aprovação <i>versus</i> Comportamento.....	16
2.2.2. Vantagens.....	17
2.2.3. Desvantagens	18
2.2.4. Uma visão moderna	19
2.3. BIG DATA	20
2.3.1. Volume.....	21
2.3.2. Variedade.....	22
2.3.3. Velocidade	22
2.3.4. Veracidade.....	23
2.3.5. Valor.....	23
2.4. PONTUAÇÃO DE CRÉDITO USANDO TECNOLOGIA BIG DATA	24
2.5. EMPRESAS QUE USAM SOLUÇÃO DE BIG DATA PARA PONTUAÇÃO DE CRÉDITO	25
3. METODOLOGIA DE PESQUISA	27
3.1. TIPO E DESCRIÇÃO GERAL DA PESQUISA	27
3.2. POPULAÇÃO E AMOSTRA	27

3.3. CARACTERIZAÇÃO DOS INSTRUMENTOS DE PESQUISA	28
3.4. ANÁLISE DE DADOS	29
3.4.1. Análise e apresentação das revistas das disciplinas	29
3.4.2. Seleção das revistas relevantes	29
3.4.3. Coleta de Dados	30
4. RESULTADOS E DISCUSSÃO	30
4.1. REVISTAS QUE MAIS PUBLICARAM	30
4.2. ANÁLISE DOS AUTORES E ARTIGOS	33
4.3. DETERMINAR LINHAS E ENFOQUES TEÓRICOS DE PESQUISA	37
4.4. ANÁLISE DE PALAVRAS CHAVE	38
4.5. ESTUDO DAS RELAÇÕES	40
5. CONCLUSÕES E RECOMENDAÇÕES	41
6. REFERÊNCIAS	43

LISTA DE FIGURAS

Figura 1 – Evolução da quantidade de trabalhos publicados	32
Figura 2 – Evolução da quantidade de citações.....	33
Figura 3 – Nuvem de palavras chaves mais recorrentes	38

LISTA DE TABELAS

Tabela 1 – Fontes classificadas segundo fator de impacto JCR.....	30
Tabela 2 – Revistas com maiores números publicações	31
Tabela 3 – Revistas com maiores números de citações	31
Tabela 4 – Artigos mais citados	34
Tabela 5 – Autores que mais publicaram	35
Tabela 6 – Países com mais publicações	36
Tabela 7 – Palavras chave mais frequentes	39
Tabela 8 – Enfoques dos principais trabalhos.....	40

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

Quando um cliente solicita o crédito, seja como tentativas de refinanciar um empréstimo ou para adquirir um imóvel, os credores em potencial muitas vezes procuram informações sobre o fundo financeiro do candidato na forma de uma pontuação de crédito fornecida por um *bureau* de crédito ou outros analistas. A pontuação de um consumidor pode influenciar a decisão do credor de estender o crédito e os termos do crédito. Em geral, os consumidores com alta pontuação são mais propensos a obter crédito e obtê-lo com melhores condições contratuais (RUSLI, 2013).

Dado que os clientes utilizam o crédito para uma série de ações que afetam a mobilidade social e financeira, tais como a compra de bens (uma casa, um automóvel), a criação de um empreendimento ou a obtenção de ensino superior; as pontuações de crédito têm um impacto considerável sobre o acesso às oportunidades, já que, até recentemente, a avaliação da confiabilidade dos consumidores se baseava unicamente na sua história financeira (WEI *et al.*, 2016). A pontuação de crédito financeiro popularizada pela *Fair Isaac Corporation* (FICO), por exemplo, conta com três dados-chave para determinar o acesso ao crédito: (a) o nível de endividamento dos consumidores, (b) a duração do histórico de crédito dos tomadores e (c) os pagamentos regulares e pontuais. Juntos, esses elementos representam cerca de 80% da pontuação FICO (WEI *et al.*, 2016).

Nos últimos anos, entretanto, a indústria de pontuação de crédito testemunhou uma mudança dramática nas fontes de dados. Um número crescente de empresas fundamenta a avaliação de solvabilidade dos consumidores em dados baseados nas mídias sociais (WEI *et al.*, 2016). Uma dessas empresas, a Lenddo, atribui classificações de crédito com base em informações nos perfis dos usuários de mídias sociais, como o histórico deles no âmbito da educação e do emprego, quantos seguidores eles têm, com quem eles têm amizades e informações sobre esses amigos (RUSLI, 2013).

Semelhante à Lenddo, um número crescente de *start ups* se especializam no uso de dados de mídias sociais e esse tópico será expandido mais a frente. Essas firmas garantem que suas práticas de *credit scoring* e financiamento de rede social podem reduzir as dúvidas dos credores sobre a contratação de candidatos com histórico financeiro pessoal limitado, que incluem muitos que são economicamente desfavorecidos, ampliando as oportunidades para uma porção maior da população e podendo beneficiar consumidores de baixa renda que de outra forma teriam dificuldade em obter crédito (WEI *et al.*, 2016).

Assim, Wei et al. (2016) mostram que há benefícios em coletar informações de uma rede de consumidores, porque quando os consumidores têm uma chance acima da média de interagir com outros de credibilidade semelhante, os laços de rede fornecem sinais confiáveis adicionais sobre a sua verdadeira solvabilidade. No entanto, é notável a falta de pesquisas que integram e sistematizam o conhecimento disponível sobre pontuação de crédito com dados de mídias sociais, ou *Big Data*, e um número crescente de especialistas consideram avanço da literatura sobre o tema e pelo crescente uso de tais práticas.

1.2. PROBLEMA DE PESQUISA

Quais as lacunas de pesquisas científicas sobre pontuação de crédito usando *Big Data*? Como tem sido a evolução do tema nos últimos anos? Quais as limitações?

1.3. OBJETIVOS

1.3.1. Objetivo Geral

Identificar a evolução de pesquisas sobre pontuação de crédito usando *Big Data*, analisando lacunas e limitações nos últimos dez anos.

1.3.2. Objetivos Específicos

- a) Conceituar Pontuação de Crédito;
- b) Conceituar *Big Data*;
- c) Explicar como o *Big Data* impacta nos modelos de pontuação de crédito;
- d) Mostrar aplicações do tema na realidade;
- e) Fazer uma revisão de literatura do tema com uso do TEMAC (Teoria do Enfoque Metaanalítico Consolidado).

1.4. JUSTIFICATIVA

Os sistemas de informação representam parte importante do capital estrutural da indústria e dos serviços 4.0 (OLIVEIRA *et al.*, 2019). Assinale-se que o capital estrutural pode ser descrito como a qualidade e o alcance dos sistemas informatizados da companhia, os seus bancos de dados e os seus conceitos organizacionais (EDVINSSON e MALONE, 1998).

Sendo assim, instituições financeiras têm usado, mediante de seus sistemas especialistas de captura, dados de pessoas que participam das mídias sociais para melhorar a percepção de risco de crédito pessoal.

No aspecto científico, a academia não tem negligenciado esse tema. Por exemplo, Onay e Öztürk (2018) sintetizaram a literatura publicada sobre Big Data e crédito pessoal entre 1976 e 2017. Porém, os enfoques de análises de revisão bibliográfica para este artigo levam para resultados diferentes dos que serão abordados neste trabalho.

Este estudo de conclusão de curso, por outra parte, usa a metodologia do TEMAC (Teoria do Enfoque Metaanalítico Consolidado), desenvolvida por Mariano et al. (2011), para resenhar a literatura aplicada a respeito de crédito pessoal e *Big Data*.

2. REFERENCIAL TEÓRICO

2.1. RISCO DE CRÉDITO

Em geral, risco é um conceito multidimensional que cobre quatro grandes grupos: risco de mercado, risco operacional, risco de crédito e risco legal (DAMODARAN, 2009). O risco de crédito, o qual importa para este estudo, está relacionado a possíveis perdas monetárias quando um dos contratantes não honra seus compromissos.

O risco de crédito é a mais antiga forma de risco no mercado financeiro. É consequência de uma transação financeira contratada entre um fornecedor de fundos (doador do crédito) e um usuário (tomador do crédito). O puro ato de emprestar uma quantia a alguém traz embutido em si a probabilidade de ela não ser recebida de volta, a incerteza em relação ao retorno. Isto é, na essência, o risco de crédito, e que se pode definir como: o risco de uma contraparte, em um acordo de concessão de crédito, não honrar seu compromisso (FIGUEIREDO, 2001).

Silva (2000), por exemplo, classifica o risco de crédito de um banco em quatro grupos: risco de administração do crédito, risco da operação, risco de concentração, e risco do cliente. O risco de administração do crédito está relacionado aos recursos humanos do banco, ou seja, quanto maior for a capacidade da instituição na avaliação do risco do cliente, menor será a probabilidade de perda. O risco da operação, por sua vez, está relacionado aos componentes de uma operação.

Os principais componentes de uma operação de crédito são: produto, montante, prazo, forma de pagamento, garantias e o preço (SILVA, 2000). A inadequação na determinação do produto ou do valor pode levar o tomador à inadimplência. O risco de uma carteira de crédito de um banco depende de sua diversificação, a regra geral

diz que quanto maior a diversificação menor será o risco da carteira. Para Caquette et al. (1999) muitas das recentes perdas em créditos sofridas por bancos resultaram de concentrações excessivas de carteiras em setores específicos. Finalmente, o risco do cliente, ou risco intrínseco, modalidade do risco de crédito bancário foco deste trabalho. O risco intrínseco compreende o tomador de crédito e decorre de suas características. Logo, nessa modalidade de risco de crédito o não cumprimento da promessa de pagamento pelo devedor decorre de um conjunto de fatores associados ao próprio devedor (SILVA, 2000).

Saunders (2000) divide as abordagens de medição do risco de crédito em tradicionais e novas. As novas abordagens seriam os modelos de gestão de carteiras, enquanto que os modelos tradicionais seriam divididos em três classes:

- (1) Sistemas especialistas;
- (2) *Credit Scoring* - sistemas de pontuação de crédito e;
- (3) Rating - sistemas de classificação.

Neste estudo o foco está no segundo modelo: pontuação de crédito, mas com uma abordagem moderna.

2.2. PONTUAÇÃO DE CRÉDITO

Pontuação de crédito pode ser definida como o processo de atribuição de pontos às variáveis de decisão mediante técnicas estatísticas. Trata-se de processo que define a probabilidade de que um cliente com certas características, pertença ou não a um grupo possuidor de outras determinadas características consideradas desejáveis, hipótese em que se aprova um limite de crédito, esta técnica, portanto, estabelece uma regra de discriminação de um determinado cliente solicitante de crédito (VICENTE, 2001).

Os sistemas de pontuação de crédito definem a probabilidade de um cliente vir a ser “bom pagador” ou “mau pagador” com base em suas características. Há vários fatores que estão associados à possibilidade de inadimplência. Um modelo de pontuação de crédito combina os fatores mais importantes associados à

possibilidade de não pagamento, determina a relação entre eles e atribui números para gerar o score final. A modelo tem por objetivo produzir um padrão de pontuação de crédito no qual quanto maior for o score, menor será o risco de perda com devedores duvidosos (GHERARDI & GHIEMETTI, 1997).

Os sistemas de pontuação de crédito são encontrados em praticamente todos os tipos de análises de crédito, desde crédito ao consumidor até empréstimos comerciais (SAUNDERS, 2000).

Apesar de o *credit scoring* representar um processo científico, ele não inibe a possibilidade de se recusar um “bom” pagador ou se aceitar um “mau” pagador. Isto ocorre porque nenhum sistema de gestão de crédito consegue o total de informações relevantes na classificação do devedor, e, mesmo que conseguisse, o seu custo tornaria a análise economicamente inviável (SOUSA & CHAIA, 2000).

2.2.1. Aprovação *versus* Comportamento

Os modelos de *credit scoring* se dividem em duas categorias: modelos de aprovação de crédito e modelos de pontuação comportamental (também conhecidos como *behavioural scoring*) (CAQUETTE et al., 1999). O *behavioural scoring* é uma ferramenta para previsão de eventos associados ao risco de crédito, como a inadimplência e pagamentos em dia, entre outras características. Para tanto leva em consideração aspectos comportamentais e de atividades de clientes da instituição (VICENTE, 2001).

A principal diferença entre as duas categorias é que nos modelos de pontuação comportamental, a instituição financeira analisa o comportamento do cliente em operações anteriores, enquanto que nos modelos de aprovação de crédito a instituição financeira não conhece o cliente. A informação adicional no *behavioural scoring* é o histórico de compras e pagamentos do cliente (THOMAS, 2000). No *behavioural scoring* são utilizadas variáveis tais como número de atrasos no último ano, volume da transação, utilização média da linha de crédito. Desta maneira cada operação recebe uma classificação, que representa seu risco de crédito relacionado ao comportamento do tomador.

Enquanto os modelos de aprovação de crédito preocupam-se apenas com a concessão e o volume de crédito, os modelos de pontuação comportamental podem ser utilizados para gestão de limites de crédito rotativo, autorizações de compra acima do limite em caso de cartões de crédito, ofertas de crédito, cobrança preventiva, entre outras estratégias (VICENTE, 2001).

2.2.2. Vantagens

Parkinson & Ochs (1998) elaboraram um resumo com as principais vantagens do uso de sistemas de *credit scoring*:

- a) Revisões de crédito consistente: os dados históricos de outros devedores são um bom indicador de consistência para revisão de crédito.
- b) Informações Organizadas: a sistematização e organização das informações contribuem para a melhoria do processo.
- c) Eficiência no trato de dados fornecidos por terceiros: o processo de *credit scoring* torna as informações de banco de dados fornecido por terceiros, anteriormente classificadas como dados acessórios, parte integrante do sistema.
- d) Diminuição da metodologia subjetiva: o uso de um sistema quantitativo parametrizado minimiza o subjetivismo.
- e) Compreensão do processo: o modelo construído sintetiza o processo de concessão de crédito da empresa, fornecendo maiores subsídios para entendê-lo.
- f) Maior eficiência do processo: a análise de crédito é centrada em um número menor de fatores, reduzindo o tempo do processo e melhorando a eficiência.

2.2.3. Desvantagens

Entre as principais limitações dos modelos de *credit scoring*, Silva (2000) destaca o aspecto temporal da amostra: o tempo (a época) é uma das principais limitações apresentadas pelos modelos desenvolvidos a partir do uso de análise discriminante. Com o decorrer do tempo, tanto as variáveis quanto seus pesos relativos sofrem alterações. As variáveis que, segundo a análise discriminante, são as que melhor classificam dentro de determinada conjuntura econômica, podem não ser em outra situação.

Para reforçar este pensamento, Caquette et al. (1999) também destacam o aspecto temporal quando afirmam que um modelo de *credit scoring* pode degradar-se pelo tempo se a população em que ele é aplicado diverge da população original que foi usada para construir o modelo. O aspecto geográfico é outra grande limitação para os sistemas de *credit scoring*, já que os aspectos da região geográfica limitam o uso de um modelo único, sendo que o desenvolvimento de diversos modelos poderia exigir que se tivesse de amostras muito grandes (SILVA, 2000).

Complementando essas limitações, Parkinson & Ochs (1998) destacaram as seguintes desvantagens do uso de sistemas de *credit scoring*: tem custo de desenvolvimento: o desenvolvimento de um sistema de *credit scoring* traz não apenas os custos do sistema, mas altos custos de manutenção com suporte técnico e recursos humanos; são modelos com “excesso de confiança”: os modelos de *credit scoring* podem adquirir rótulos de “perfeitos”, não permitindo crítica de seus resultados; tem problemas de “valores não preenchidos”: a falta de dados não característicos de determinados clientes, ou mesmo a falta de informação pode causar problemas na sua utilização, assim como resultados não esperados; uma possível interpretação equivocada dos escores com o uso inadequado do sistema, devido à falta de conhecimento pode ocasionar problemas para a instituição.

2.2.4. Uma visão moderna

Sendo assim, é notável que conceitos de pontuação de crédito mudassem em mais de vinte anos, sendo alguns defasados e outros considerados futuristas, em ambos os casos levando em consideração o contexto em que foram construídos. Pode-se pontuar que na crítica de Souza e Chaia (2000) ao modelo de *credit scoring* (em que falam sobre nenhum sistema de gestão de crédito conseguir o total de informações relevantes na classificação do devedor, pois mesmo que conseguisse o seu custo tornaria a análise economicamente inviável) há uma defasagem de contexto temporal, porque talvez naquela época fosse inimaginável ter um sistema que conseguisse gerar informações precisas e relevantes, derivadas de um grande volume de dados. Com o surgimento de diversas tecnologias para análises de sistemas de informação e de dados em grande volume, esse processo de gerar mais e mais informações precisas e relevantes rapidamente tem se tornado mais barato e, conseqüentemente, economicamente viável para empresas estarem usando da melhor solução para avaliar seus clientes de modo rápido e eficaz.

Outro ponto que vale comentar é sobre a definição de *behavioural scoring* de Thomas (2000) e Caquette et al. (1999) – na qual falam que a informação adicional entre esse modelo de crédito em relação aos outros é o histórico de compras e pagamentos do cliente – que está ultrapassada no que tange sua composição, isto é, quando se fala de pontuação comportamental hoje se engloba outras informações além do histórico de transações financeiras; mas também informações de ligações telefônicas, publicações em redes e mídias sociais, trajetos marcados por GPS via *smartphones* e entre outras fontes de dados não estruturados como será citado mais adiante. Assim como é explicado por Neoway (2019) o *Big Data Analytics* funciona em duas pontas principais, pois permitem reunir em um único sistema dados estruturados e atualizados de diferentes fontes, e possibilitam o cruzamento rápido destas informações. Assim, é possível ter uma visão mais ampla e em tempo real dos riscos relacionados a cada perfil de cliente e realizar análises preditivas para a tomada precisa de decisão (NEOWAY, 2019).

É possível capturar e avaliar dados de fontes externas como SPC, Serasa, Receita Federal, mídias sociais, ligações telefônicas, rastro de cliques entre sites e aprofundar as análises de acordo com os modelos e métricas relevantes (combinando diferentes critérios de busca) para cada empresa. Este tipo de combinação não seria possível com os processos tradicionais (e muitas vezes manuais) de confirmação de dados. Pode-se então, com escala e velocidade, montar um estudo mais completo do potencial cliente (NEOWAY, 2019).

Por fim, destacando o pensamento visionário que Silva (2000) coloca de modo a confirmar que a tecnologia neste campo de estudo veio para tornar possíveis modelos de avaliação de risco mais rápidos e precisos. Silva (2000) ressalta que quanto mais agilidade as instituições financeiras que concedem crédito conseguem ganhar, mais poderão dedicar seus tempos a outros assuntos relevantes e que não possam ser sistematizados, ao invés do analista ficar examinando e concluindo sobre cada um dos índices.

2.3. BIG DATA

Big Data é uma tecnologia que permite o processamento de informações com alto desempenho e disponibilidade. Um conjunto de ferramentas digitais que tornam a coleta, o processamento e a visualização de dados mais simples, padronizadas e eficazes. Desse modo, os gestores podem entender com mais clareza as tendências e os padrões para organizar sua estratégia de negócios (NEOWAY, 2019).

O termo *Big Data* é utilizado para definir um grande conjunto de ferramentas de Tecnologia da Informação (TI) que permitem a captura, a análise e a catalogação de registros em tempo real. As informações podem ser originadas de diferentes fontes internas e externas, como cadastro de clientes, análises de mercado, mídias sociais, dispositivos eletrônicos, processos internos ou mesmo pesquisas em meios *offline* (NEOWAY, 2019).

A vantagem dessas ferramentas está em centralizar, em um único local, a coleta e a análise desse grande conjunto de registros. A partir disso, as técnicas de estatística

e processamento ficam a cargo das máquinas, permitindo que analistas consigam identificar padrões rapidamente e prever tendências com maior precisão. Como consequência, será possível criar rotinas mais eficazes e se preparar para as mudanças do mercado antecipadamente, isto é o que se nomeia como análises preditivas. Assim, o negócio pode manter-se continuamente com um alto grau de competitividade (NEOWAY, 2019).

O conceito completo de *Big Data* está baseado em cinco princípios e cada um deles terá influência direta no desempenho das soluções disponibilizadas no mercado. Esses princípios são conhecidos como os 5 V's do *Big Data*. Para Neoway (2019) existem conteúdos que apontam até 10 V's, mas nesse estudo adota-se que são os cinco principais que originam os demais, são eles: volume, variedade, velocidade, veracidade e valor.

2.3.1. Volume

Uma ferramenta de Big Data deve ser capaz de lidar com um grande volume de dados. É mediante às mídias sociais, smartphones, internet móvel e os dispositivos conectados por meio da Internet das Coisas (IoT), a quantidade de informações que circula em meios digitais cresce continuamente. A tendência é que o volume cresça a cada minuto entre postagens no Twitter, Facebook e Instagram, mensagens de e-mail, aplicativos de conversa, entre outros tipos de arquivos que circulam na 'nuvem' de servidores mundiais (NEOWAY, 2019).

Portanto, cada vez mais se estabelece uma dependência de ferramentas de *Big Data*, que mediante Inteligência Artificial e aprendizado de máquinas chegou-se a um novo padrão de análise de dados. Essas tecnologias permitem que analistas consigam trabalhar com um grande fluxo de dados com alto desempenho - muitas vezes, as informações são criadas e coletadas em tempo real. Assim, os sistemas de *Big Data* devem conseguir lidar com tais fluxos de informações sem gerar perda de desempenho ou alto custo computacional (NEOWAY, 2019).

2.3.2. Variedade

Uma solução de *Big Data* tem capacidade de trabalhar com fluxos de dados variados. As informações podem ser originadas em equipamentos diversos, mídias sociais, dispositivos móveis e mesmo em meios *offline*, como pesquisas de mercado e tabelas com dados de transações financeiras. Assim, cada dado tem um tipo de característica e, por isso, são classificados em dois tipos: os dados não estruturados e estruturados (NEOWAY, 2019).

Os dados não estruturados são os que não são coletados de um banco de dados ou os que não têm uma estrutura definida, por que são registros que devem ser catalogados antes de analisados, uma vez que têm um número maior de ruídos (informações sem relevância para o analista). Esse é o caso, por exemplo, de mensagens de texto, e-mails, documentos do *Word*, apresentações de *Powerpoint* e mesmo arquivos de mídia (áudio e vídeo) (NEOWAY, 2019).

Por outro lado, os dados estruturados são os que, quando coletados, já estão organizados em um banco de dados ou solução semelhante. São mais fáceis de serem avaliados por ferramentas de *Big Data*, uma vez que o número de rotinas que devem ser executadas para a avaliação do conteúdo é menor (NEOWAY, 2019).

Portanto, em relação à variedade, uma solução de *Big Data* deve ser capaz de lidar com ambos os tipos de conteúdos. Caso ela seja mal programada, o custo computacional para que analistas executem o seu trabalho será alto. Além disso, a confiabilidade dos *insights* obtidos será reduzida (NEOWAY, 2019).

2.3.3. Velocidade

Com o fluxo contínuo de dados em grande quantidade, uma solução de *Big Data* precisa ter um alto desempenho de análise, de tal forma que seja possível encontrar padrões rapidamente. Diante disso, tecnologias auxiliares são utilizadas para garantir o maior desempenho de soluções de *Big Data* (NEOWAY, 2019).

A computação na nuvem, por exemplo, é uma das principais “aliadas” do *Big Data*. Ao executar tais sistemas na nuvem, analistas ganham maior escalabilidade operacional com um custo menor. Assim, caso o fluxo de informações aumente, é possível escalar os recursos, impedindo que a nova demanda impacte na velocidade de execução das rotinas de análise (NEOWAY, 2019)

2.3.4. Veracidade

É crucial conjuntos de dados confiáveis. Como já citado, muitas vezes os registros utilizados não são estruturados, o que pode levar cenários em que o número de ruídos é alto, impactando na qualidade do trabalho do analista (NEOWAY, 2019).

Diante disso, as soluções de *Big Data* devem estar preparadas para buscar por dados de fontes confiáveis, e devem dar a possibilidade de filtrar quais conteúdos são relevantes para o negócio, além de eliminar aqueles não são confiáveis ou que não têm relevância. Assim, cria-se uma rotina de análise mais precisa e com maior chance de sucesso (NEOWAY, 2019).

Vale ressaltar que a segurança de informações é outro aspecto essencial, especialmente devido às novas regulações de Proteção de Dados. Assim, é fundamental a manutenção da integridade da informação para o armazenamento, compartilhamento e análise, evitando vazamentos e usos indevidos previstos pelas legislações nacional e internacional (NEOWAY, 2019).

2.3.5. Valor

Afinal, para entender o que é *Big Data*, tem-se o aspecto do valor. Ou seja, a solução deve ser capaz de agregar valor a processos e tornar serviços mais competitivos. Identificar tendências e padrões que gerem possibilidades para o gestor tomar decisões com confiança e direcionar melhor a estratégia para conquistar clientes e/ou mais mercados. Ou ainda, pelo aspecto operacional, poderá avaliar as rotinas internas e uso de ferramentas corporativas para rastrear gargalos e tornar a gestão de processos mais eficiente (NEOWAY, 2019).

2.4. PONTUAÇÃO DE CRÉDITO USANDO TECNOLOGIA BIG DATA

Percebe-se que o *Big Data* tem várias áreas de impacto e neste estudo será restringido ao impacto da tecnologia nas análises de risco e pontuação de crédito.

A gestão de riscos eficiente para os processos de concessão de crédito deve reunir informações relevantes, do maior número possível de fontes, que devem estar acessíveis em tempo integral, serem constantemente atualizadas e facilmente compreendidas por todos os envolvidos no processo (NEOWAY, 2019).

Para um mínimo de segurança ao conceder crédito, deve-se considerar uma série de informações do futuro cliente, relacionadas ao seu CNPJ ou CPF. Estes dados são agrupados para formar a pontuação de crédito, em que se determina um perfil do futuro cliente: se ele pode ser um bom pagador e com boas condições financeiras para quitar suas dívidas, ainda mais quando se tratam de operações envolvendo grandes valores (NEOWAY, 2019).

Entretanto, existem dois pontos fundamentais neste processo de avaliação do perfil do cliente: precisão e agilidade. Mas como determinar exatamente (ou com a maior certeza possível) para que tipo de pessoa, ou empresa, será feita a concessão de crédito sem demorar dias para validar o empréstimo? A resposta está no *Big Data Analytics* e é sobre como a tecnologia aprimora a análise de dados de tanto de pessoas físicas e jurídicas (NEOWAY, 2019).

Sendo assim, a tecnologia garante operação eficiente, rápida e segura na concessão de crédito. As empresas que atuam com concessão de crédito têm foco voltado para prevenção, buscando evitar perdas financeiras com inadimplências. Assim, adotam procedimentos que possam ajudar a reduzir este risco. Por isso, devem analisar com precisão o perfil do cliente. Mas também devem estar atentos ao tempo de duração desta avaliação, a fim de evitar a perda do negócio, pois o concorrente pode estar um passo à frente e conseguir fechar um empréstimo rentável antes de uma empresa que não é rápida, porque conseguem verificar

rapidamente um grande volume de dados, determinando com precisão e agilidade um score de PF ou PJ (NEOWAY, 2019).

2.5. EMPRESAS QUE USAM SOLUÇÃO DE BIG DATA PARA PONTUAÇÃO DE CRÉDITO

Com a modernidade novas empresas surgem no mercado com soluções inovadoras e competitivas para tornar as decisões financeiras mais ágeis, precisas e inclusivas quando as instituições financeiras não estavam fazendo isso. É bem interessante como a tecnologia veio para causar movimentos fortes de melhoria na economia e no mercado financeiro, porque como os bancos tradicionais estão perdendo espaço para bancos digitais e *'fintechs'*, para sobreviver a todas as mudanças, essas instituições são forçadas a acompanhar o ritmo e conseqüentemente adotar as mesmas tecnologias ou ainda melhores.

Quando se fala de inclusão mediante pontuação de crédito com dados alternativos usando soluções de *Big Data*, não necessariamente se diz respeito a uma concessão de crédito relativamente alto para alguém que não pode nem mesmo honrar aquele compromisso só porque devido ao modelo de análise está pessoa é confiável. Pode ser concedido mil reais em limite para um universitário sem renda fixa declarada, mas, em termos absolutos, mil reais de crédito é alto para quem não tem histórico de transação e nem renda fixa. Sendo assim, é relevante mostrar que o objeto deste estudo tem relevância que vai além de abordagens acadêmicas, uma vez que existe na realidade e impacta de modo significativo na vida de pessoas, bancos e empresas.

A 'Tesla ADS' é uma empresa que através de redes neurais consegue fazer análises de crédito e risco em tempo real para detectar fraudes. Segundo a empresa, a precisão é de 94% para um tempo de resposta de apenas 1 segundo. Tem a 'Easy Crédito' que mediante uso de aplicativo faz análise de crédito para compradores, o usuário diz o que deseja comprar e onde e o sistema mostra se ele tem crédito aprovado para realizar a compra; se a loja selecionada não estiver disponível, é possível visualizar outras marcas que oferecem crédito (ZUINI, 2015).

A Lenddo, Neo Finance e Affirm também fazem parte de um número crescente de empresas de crédito que usam dados pessoais tirados de sites de mídias sociais como o Facebook, LinkedIn e Twitter para avaliar o risco de crédito do consumidor, conforme diz Rusli (2013). Elas acreditam que a situação social da pessoa, sua reputação online e/ou conexões profissionais são fatores a serem levados em conta na hora de conceder crédito, principalmente para alguém com histórico de crédito irregular ou escasso e quem, de outro modo, teria dificuldade em conseguir um empréstimo (WHARTON UNIVERSITY OF PENNSYLVANIA, 2013).

A Neo Finance de Palo Alto, na Califórnia tem como público alvo tomadores com histórico de crédito recente e que seriam candidatos certos a empréstimos com taxas de juros elevadas por parte dos credores tradicionais. A empresa oferece empréstimos a taxas de juros mais baixas depois de analisar informações como o histórico de emprego do candidato e a qualidade das conexões do seu perfil no LinkedIn, aferindo assim o potencial de ganhos futuros e a estabilidade no emprego da pessoa. Ela leva em conta a pontuação FICO (modelo criado pela Fair & Isaac Corporation), mas só para se inteirar de possíveis suspeitas (WHARTON UNIVERSITY OF PENNSYLVANIA, 2013).

A Lenddo, de Hong Kong, vai além e usa as conexões sociais do devedor para pressioná-lo caso ele deixe de fazer um pagamento (RUSLI, 2013). Por exemplo, a start up informará seus amigos do Facebook caso ele deixe de fazer um pagamento. Com isso, a pontuação dessas pessoas na Lenddo será prejudicada se o cliente não pagar o que deve. A empresa calcula sua pontuação de crédito de 1 a 1.000 depois de avaliar 100 bancos de dados e mídias sociais em busca de informações como a localização do candidato e o número de conexões que ele tem (WHARTON UNIVERSITY OF PENNSYLVANIA, 2013).

Por fim, no mesmo sentido de análise da Lenddo, a Affirm (*start up* de São Francisco administrada por Max Levchin, um dos fundadores da PayPal) facilita o pagamento de bens e serviços através do smartphone — com dois toques na tela — e dá ao cliente 30 dias para acertar suas contas sem cobrança de multa (RUSLI, 2013).

3. METODOLOGIA DE PESQUISA

A metodologia adotada foi uma revisão sistemática da bibliografia, de carácter exploratório por meio do TEMAC (Teoria do enfoque metaanalítico consolidado). O enfoque utiliza o critério de impacto de revistas e artigos para escolha do material a ser utilizado (CALAZANS et al., 2015).

3.1. TIPO E DESCRIÇÃO GERAL DA PESQUISA

O tipo de pesquisa é uma revisão sistemática da bibliografia, de carácter exploratório com a abordagem do TEMAC, o qual utiliza da revisão qualitativa, integrativa e sistemática, podendo em análises mais profundas utilizar o metaanálises como uma análise final.

3.2. POPULAÇÃO E AMOSTRA

Todos os dados foram extraídos pelas bases de dados *Web of Science*TM e *Journal Citation Reports*. A pontuação de crédito usando de *Big Data* é uma área de conhecimento que tem interface com várias outras. Sendo assim, escolheu-se as quatro categorias (*Business, Management, Economics* e *Business Finance*) que mais publicaram artigos vinculados ao assunto estudado.

Após a identificação das revistas com maior fator de impacto (com fator de impacto acima de 9) foi realizada a busca com os termos “*credit scoring*” AND “*Big Data*” (pontuação de crédito E super dados), no espaço temporal de dez anos (2009 a 2018), e dentro das quatro categorias que mais publicaram artigos relacionados ao assunto estudado. A busca no *Web of Science* resultou em vinte e quatro textos, entre artigos, reviews e proceedings papers.

3.3. CARACTERIZAÇÃO DOS INSTRUMENTOS DE PESQUISA

O instrumento de pesquisa usado foi o TEMC. Basicamente, a TEMAC tem por objetivo unir o que é utilizado no enfoque meta analítico garantindo a qualidade na avaliação. Dessa forma a análise foi realizada segundo a TEMAC, que é um modelo de revisão sistemática dividida em sete passos, explicadas por Mariano, Garcia Cruz e Arenas Gaitán (2011). Os passos são mostrados resumidamente a seguir:

- a) Análise e apresentação das revistas da disciplina: conhecer as revistas científicas mais utilizadas para a busca do tema;
- b) Seleção das revistas relevantes: selecionar as revistas mais relevantes para o tema estudado;
- c) Coleta de dados: realizar pesquisa sobre o tema em bases de dados;
- d) Análise dos autores e artigos: busca-se saber a relevância do tema ao longo do tempo, e numa segunda análise quais são os autores mais importantes no período escolhido da pesquisa;
- e) Determinação dos enfoques teóricos de pesquisa: estabelecer quais as principais linhas de pesquisa que abordam o tema;
- f) Análise das palavras-chave: fornece elementos a respeito da evolução do tema e quais as linhas de pesquisa dentro dele;
- g) Estudo das relações: identificar o tamanho das amostras dos estudos, técnicas estatísticas utilizadas, locais de estudo, e signo e significância das relações.

3.4. ANÁLISE DE DADOS

Nas subseções seguintes, detalham-se as etapas seguidas pela pesquisa com enfoque metaanalítico e os resultados encontrados. Todos os dados foram extraídos pelas bases de dados Web of Science™ e Journal Citation Reports. As buscas ocorreram entre 09/09/2019 e 16/09/2019. O Web of Science™ é hoje uma plataforma de pesquisa, ajudando o pesquisador a rapidamente encontrar, analisar, e compartilhar informações sobre ciências, ciências sociais, artes e humanidades (CLARIVATE ANALYTICS, 2019). O InCites™ Journal Citation Reports é uma base onde pode-se encontrar e categorizar as revistas de diversas disciplinas.

3.4.1. Análise e apresentação das revistas das disciplinas

A pontuação de crédito usando de *Big Data* é uma área de conhecimento que tem interface com várias outras. A fim de priorizar os principais domínios de conhecimento relacionados à pontuação de crédito, escolheu-se as quatro áreas de pesquisa que mais foram publicados artigos vinculados ao assunto estudado. Elas são: *Business, Management, Economics e Business Finance*. Foram encontradas 688 revistas na base do *Journal Citation Reports* (JCR).

3.4.2. Seleção das revistas relevantes

O fator de impacto maior que nove foi o critério escolhido para listar as principais fontes, abaixo desse fator há muitas revistas com notas semelhantes, ficando na área de fator comum e não diferenciado como as revistas que estão acima de nove, que são poucas. O fator de impacto de um periódico acadêmico é uma medida que reflete o número médio de citações nos artigos recentes publicados no periódico (INSTITUTE FOR SCIENCE INFORMATION, 2014). As cinco revistas/conferências mais relevantes de um total de seiscentas e oitenta e oito registros, dentro das quatro áreas analisadas, são apresentadas na Tabela 1.

Tabela 1 – Fontes classificadas segundo fator de impacto JCR

Rank	Título da Revista/Conferência	Fator de Impacto
1	ACADEMY OF MANAGEMENT ANNALS	12,289
2	QUARTERLY JOURNAL OF ECONOMICS	11,775
3	ACADEMY OF MANAGEMENT REVIEW	10,632
4	JOURNAL OF THE ACADEMY OF MARKETING SCIENCE	9,360
5	JOURNAL OF MANAGEMENT	9,056

3.4.3. Coleta de Dados

Após a identificação das revistas com maior fator de impacto, foi realizada a busca com os termos “*credit scoring*” AND “*Big Data*” (‘pontuação de crédito’ E ‘Big Data’), no espaço temporal de dez anos (2009 a 2018), e dentro das quatro categorias que mais publicaram artigos relacionados ao assunto estudado. A busca no ‘*Web of Science*’ resultou em vinte e quatro textos, entre artigos, reviews e proceeding papers.

4. RESULTADOS E DISCUSSÃO

Os resultados mostrados a seguir levaram em consideração o passo 2 do TEMAC.

4.1. REVISTAS QUE MAIS PUBLICARAM

Houve vinte e quatro revistas/conferências que publicaram sobre o tema. Como houve uma pulverização grande dos resultados, foram evidenciadas as vinte e quatro revistas que publicaram, já que o resultado aponta que cada fonte publicou apenas uma vez sobre o assunto.

Tabela 2 – Revistas com maiores números publicações

Título da Revista	Registros	% de 24
2017 4TH INTERNATIONAL CONFERENCE ON INDUSTRIAL ECONOMICS SYSTEM AND INDUSTRIAL SECURITY ENGINEERING IEIS	1	4,167
AMERICAN ECONOMIC JOURNAL APPLIED ECONOMICS	1	4,167
BUSINESS INFORMATION SYSTEMS BIS 2016	1	4,167
COMPETITION CHANGE	1	4,167
ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS	1	4,167
ENTERPRISE AND COMPETITIVE ENVIRONMENT	1	4,167
INTERNATIONAL JOURNAL OF FINANCIAL ENGINEERING	1	4,167
JOURNAL OF BEHAVIORAL AND EXPERIMENTAL ECONOMICS	1	4,167
JOURNAL OF BUSINESS RESEARCH	1	4,167
JOURNAL OF DEVELOPMENT STUDIES	1	4,167
JOURNAL OF FINANCIAL REGULATION AND COMPLIANCE	1	4,167
LECTURE NOTES IN BUSINESS INFORMATION PROCESSING	1	4,167
MANAGEMENT DECISION	1	4,167
MATHEMATICAL METHODS IN ECONOMICS MME 2014	1	4,167
NEW POLITICAL ECONOMY	1	4,167
PRAGUE ECONOMIC PAPERS	1	4,167
PROCEEDINGS 31ST EUROPEAN CONFERENCE ON MODELLING AND SIMULATION ECMS 2017	1	4,167
PROCEEDINGS OF THE 10TH EUROPEAN CONFERENCE ON KNOWLEDGE MANAGEMENT VOLS 1 AND 2	1	4,167
PROCEEDINGS OF THE 8TH EUROPEAN CONFERENCE ON INTELLECTUAL CAPITAL ECIC 2016	1	4,167
PROCEEDINGS OF THE EUROPEAN CONFERENCE ON INTELLECTUAL CAPITAL	1	4,167
QUANTITATIVE FINANCE	1	4,167
QUARTERLY REVIEW OF ECONOMICS AND FINANCE	1	4,167
REVIEW OF INTERNATIONAL POLITICAL ECONOMY	1	4,167
SERBIAN JOURNAL OF MANAGEMENT	1	4,167
TOTAL	24	100,00

E as que obtiveram mais citações foram listadas na Tabela 3.

Tabela 3 – Revistas com maiores números de citações

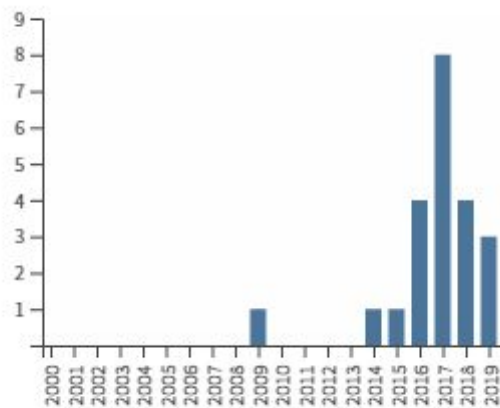
Título da Revista	Ano de Publicação	Total de Citações
JOURNAL OF BEHAVIORAL AND EXPERIMENTAL ECONOMICS	2016	13
COMPETITION & CHANGE	2017	10
NEW POLITICAL ECONOMY	2017	10
JOURNAL OF DEVELOPMENT STUDIES	2015	7
AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS	2016	6
ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS	2018	4

O que pode ser percebido é que apesar de o assunto estar bastante disperso em vinte e quatro fontes de publicações, há uma concentração de estudos relevantes em apenas seis revistas, já que no restante das fontes os estudos têm apenas uma citação ou nenhuma. Mas vale lembrar que muitos destes estudos foram publicados há menos de dois anos e o mais citado foi publicado em 2016, há mais de dois anos. Logo, pode-se dizer que em consideração a “juventude” do tema essa concentração de alguma relevância em apenas seis das vinte e quatro revistas sendo que cada uma das revistas publicou uma vez, isso não quer dizer necessariamente que os outros estudos não sejam relevantes, mas sim recentes e não houve tempo para que pudessem ser citados ainda.

De qualquer forma, nenhuma das revistas mostradas nas duas tabelas anteriores (tabelas 2 e 3) está entre as que possuem maiores fatores de impacto nas áreas de interesse desse estudo.

Sendo assim, para complementar a análise, se fez uma evolução temporal das discussões a respeito do tema, a seguir apresenta-se a quantidade de publicações (Figura 1) e o número de citações feitas nos últimos dez anos (Figura 2):

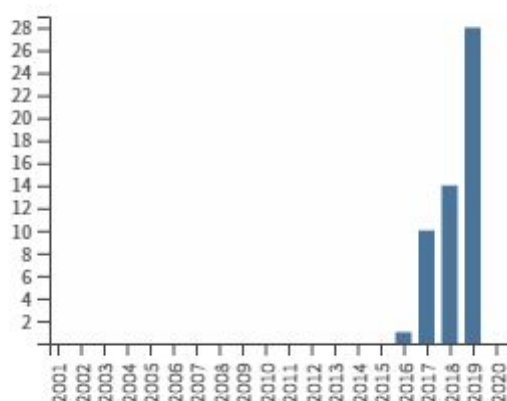
Figura 1 – Evolução da quantidade de trabalhos publicados



A Figura 1 mostra que a primeira publicação do assunto, dentro do período pesquisado, em uma revista de relevância acadêmica foi em 2009, tendo uma carência de quatro anos (entre 2010 e 2013) para publicação de novos estudos. Entretanto, a partir de 2014, houve grande crescimento do número de publicações, chegando ao pico em 2017, devido a uma popularização intensa do assunto, e posteriormente houve uma queda, já que para estudar sobre este tema é difícil uma vez que poucos são os que têm recursos para usar de *Big Data* em seus processos, logo se não está se utilizando na realidade, não tem pesquisa.

O número de citações tem crescido no decorrer do tempo, apesar de o tema ter tido publicações anteriores a 2016 é notável que só a partir desse ano que as publicações sobre o assunto começaram a ser citadas.

Figura 2 – Evolução da quantidade de citações



O alcance representativo é crescente, comprovado por meio da evolução das citações ao longo dos anos representado pela Figura 2. A soma do número de citações sobre esse assunto até o ano de 2019 chega a cinquenta e três, sendo o mais citado com treze e bem acima da média de 2,5 por ano.

Em termos absolutos, pode-se inferir que o interesse acadêmico e a relevância do tema são ainda baixos quando em comparação com outros assuntos. Contudo, existe alto crescimento e interesse em pesquisas sobre modelos de pontuação de crédito usando *Big Data*, levando em consideração o comportamento do gráfico no decorrer do tempo.

4.2. ANÁLISE DOS AUTORES E ARTIGOS

Para a análise dos autores, buscou-se inicialmente estabelecer uma relação entre os textos mais citados e os autores com maior número de textos publicados. A Tabela 4 mostra os artigos que tiveram mais citações no decorrer dos últimos dez anos:

Tabela 4 – Artigos mais citados

Autores	Título	Citações	Ano de Publicação
Rustichini, Aldo; DeYoung, Colin G.; Anderson, Jon E.; Burks, Stephen V.	Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation	13	2016
Aitken, Rob	All data is credit data: Constituting the unbanked	10	2017
Campbell-Verduyn, Malcolm; Goguen, Marcel; Porter, Tony	Big Data and algorithmic governance: the case of financial practices	10	2017
Kislat, Carmen	Why are Informal Loans Still a Big Deal? Evidence from North-east Thailand	7	2015
Einav, Liran; Finkelstein, Amy; Kluender, Raymond; Schrimpf, Paul	Beyond Statistics: The Economic Content of Risk Scores	6	2016
Xia, Yufei; Yang, Xiaoli; Zhang, Yeying	A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending	4	2018

A Tabela 5 mostra os autores que mais publicaram e o número de citações que obtiveram com os textos que publicaram na área.

Tabela 5 – Autores que mais publicaram

Autores	Publicações	Citações
ANDERSON JE	1	13
BURKS SV	1	13
DEYOUNG CG	1	13
RUSTICHINI A	1	13
AITKEN R	1	10
CAMPBELL-VERDUYN M	1	10
GOGUEN M	1	10
PORTER T	1	10
KISLAT C	1	7
EINAV L	1	6
FINKELSTEIN A	1	6
KLUENDER R	1	6
SCHRIMPF P	1	6
XIA YF	1	4
YANG XL	1	4
ZHANG YY	1	4
BERNARDS N	1	1
DEMYANYK Y	1	1
MALGIER G	1	1
AL-SHAWABKEH A	1	0
BEKIROS S	1	0
BOUDRIGA N	1	0
DJEMAIEL Y	1	0
HA HY	1	0
KALINA J	1	0
KANG MS	1	0
KUC M	1	0
LABIDI N	1	0
LAHMIRI S	1	0
MADAR L	1	0
ONAY C	1	0
OZTURK E	1	0
PAN HF	1	0
PEREZ-MARTIN A	1	0
PEREZ-TORREGROSA A	1	0
RYCHNOVSKY M	1	0
SOLDATYUK N	1	0
SOPKO S	1	0
TAMBYRAJAH A	1	0
TEPLY P	1	0
VACA M	1	0
WANG HQ	1	0
YAN HY	1	0
YUAN GXZ	1	0

É notável que as publicações, na amostra, estão dispersas não somente em sua fonte, mas também por quem as publicou. De vinte e quatro artigos, há quarenta e quatro autores listados e nenhum publicou mais de uma vez, isso quer dizer que todos os artigos da amostra foram publicados por autores diferentes, os autores que escreveram os artigos mais citados escreveram apenas aqueles. Outro fato é que os artigos mais citados foram anexados à base de dados na metade do período pesquisado, 2016. Pode-se inferir que as publicações de tal período estão servindo de base para as mais recentes.

Finalmente, apresenta-se nessa subseção, na Tabela 6, a distribuição de textos por país. Há registro de catorze países com artigos publicados sobre o tema. A República Tcheca e China lideram, e ambos os países tem o mesmo número de artigos registrados. No entanto, é notável que exista uma dispersão forte entre os países de publicação, porque a diferença entre os primeiros colocados e os últimos colocados é de no máximo três estudos. Isso mostra que todos os países listados tem número de publicações muito próximo e igual. O Brasil não publicou nada sobre esse tema.

Tabela 6 – Países com mais publicações

Países	Registros	% of 24
CZECH REPUBLIC	4	18.182
PEOPLES R CHINA	4	18.182
CANADA	3	13.636
ENGLAND	3	13.636
USA	3	13.636
GERMANY	2	9.091
ITALY	2	9.091
GREECE	1	4.545
HUNGARY	1	4.545
MOROCCO	1	4.545
SOUTH KOREA	1	4.545
SPAIN	1	4.545
TUNISIA	1	4.545
TURKEY	1	4.545

4.3. DETERMINAR LINHAS E ENFOQUES TEÓRICOS DE PESQUISA

Para determinar as principais linhas de pesquisa, foram analisados os principais artigos relacionados ao tema. Concluiu-se que a pontuação de crédito com uso de *Big Data* está bastante vinculada às decisões financeiras, *internet banking* (bancos online), modelos de risco, redes, privacidade de informação e análises de comportamento. A Figura 3 mostra os principais enfoques de modo visual, com o uso da ferramenta disponível no endereço eletrônico TagCrowd.com. A ferramenta analisou o conteúdo dos vinte e quatro artigos, mostrando as palavras mais recorrentes em tamanho maior. Foram excluídas as palavras “credit”, “scoring” “big” e “data”, por se tratarem de palavras utilizadas na busca e que, conseqüentemente, apareceriam em maior tamanho.

Figura 3 – Nuvem de palavras chaves mais recorrentes



4.4. ANÁLISE DE PALAVRAS CHAVE

Para identificar as palavras chave que podem guiar novas pesquisas, são apresentadas, na Tabela 7, as dez mais recorrentes em todos os textos encontrados:

Tabela 7 – Palavras chave mais frequentes

Palavra Chave	Frequência
Risk	11
Classification	7
Model	7
Financial	6
Networks	5
Banking	4
Finance	4
Privacy	4
Inclusion	3
Trade	3

Analisando o contexto em que tais palavras foram empregadas, foi observado que os artigos usam, criticam e propõem modelos e/ou sistemas de estimativa de risco, controle, classificação para pontuação de crédito e análise de fontes de dados alternativos para os mais diversos objetivos, tais como: aprimorar modelos de pontuação de crédito; inclusão social de pessoas que poderiam ser potenciais clientes, mas que não tem histórico de transação bancária; criticar modelos de desiguais de avaliação de crédito de *'fintechs'*, como lidar com a privacidade dos dados na era do big data; como a rede social de contatos influência na solvabilidade de um cliente etc.

Também se percebeu que havia três grupos de artigos escritos ao longo dos últimos dez anos e que foram predominantes:

- a) Os que revisam e propõem modelos de pontuação de crédito para facilitar o acesso de pessoas físicas e jurídicas usando soluções de big data e, por outro lado, aprimorar as decisões de instituições financeiras na hora de conceder o crédito.
- b) Os que abordam revisões sobre literatura de risco de crédito na era do big data e buscam analisar a evolução do tema.
- c) E o que tratam de análises do desempenho de modelos de avaliação de crédito utilizados, propondo melhorias.

4.5. ESTUDO DAS RELAÇÕES

O último passo consiste em conhecer melhor os enfoques e aspectos metodológicos das pesquisas realizadas, mostrados na Tabela 8. Escolheram-se apenas os artigos com mais citações, já que todos os autores publicaram a mesma quantidade.

Tabela 8 – Enfoques dos principais trabalhos

Autores	Título	Enfoque/Amostra	Citações
Rustichini, Aldo; DeYoung, Colin G.; Anderson, Jon E.; Burks, Stephen V.	Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation	O objetivo é uma avaliação sistemática do padrão empírico de relacionamentos e correlações entre as medidas usadas na economia e psicologia da personalidade. Os resultados mostram que os traços de personalidade têm um poder preditivo estatístico comparável ou mais forte do que as preferências econômicas para várias variáveis dependentes, incluindo pontuação de crédito, persistência no emprego e acidentes com caminhões pesados. Amostra: caminhoneiros trainees em uma grande empresa dos EUA.	13
Aitken, Rob	All data is credit data': Constituting the unbanked	Este artigo examina um conjunto de novas práticas projetadas para tornar visível - e extrair valor daqueles sem pontuação formal de crédito nos mercados financeiros contemporâneos. Muitos projetos de inclusão financeira agora tentam classificar o crédito como invisível, baseando-se em uma variedade de dados alternativos e por avaliar esses dados em relação aos modelos de avaliação de risco com base na análise de big data.	10
Campbell-Verduyn, Malcolm; Goguen, Marcel; Porter, Tony	Big Data and algorithmic governance: the case of financial practices	Este artigo contrasta visões utópicas ou distópicas sobre a crescente presença do Big Data nas práticas financeiras contemporâneas.	10
Kislat, Carmen	Why are Informal Loans Still a Big Deal? Evidence from North-east Thailand	Este artigo examina o uso e o benefício de empréstimos informais para diferentes grupos de renda de famílias rurais no nordeste da Tailândia.	7
Einav, Liran; Finkelstein, Amy; Kluender, Raymond; Schrimpf, Paul	Beyond Statistics: The Economic Content of Risk Scores	Neste artigo, é observado que esses escores estatísticos amplamente utilizados resumem uma heterogeneidade muito mais rica e podem ser endógenos ao contexto em que são aplicados. Amostra: dados do Medicare Parte D	6
Xia, Yufei; Yang, Xiaoli; Zhang, Yeying	A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending	Este estudo aborda o problema da inferência de rejeição em um domínio específico de empréstimos P2P da perspectiva da aprendizagem semi supervisionada. Um novo método de inferência de rejeição (CPLELightGBM) é proposto combinando a estrutura contrastante de estimativa de verossimilhança pessimista e um classificador avançado de árvore de decisão (LightGBM).	4

Os objetivos dos seis artigos apresentados são diversificados. Apenas um estudo propõe novas contribuições teóricas com novos modelos e metodologias (quatro citações). Há três que usam conhecimentos passados e ferramentas para obter

conclusões a respeito de casos específicos (vinte e seis citações). Os outros dois estudam como novos conceitos sobre risco de crédito usando *big data* estão sendo aplicados na prática (vinte citações).

5. CONCLUSÕES E RECOMENDAÇÕES

O estudo buscou identificar a evolução de pesquisas sobre pontuação de crédito usando *Big Data*, analisando lacunas e limitações nos últimos dez anos (2009 – 2018), utilizando o enfoque metaanalítico consolidado como ferramenta de metodologia para produção do trabalho.

A metodologia foi uma pesquisa bibliográfica de caráter exploratório, com enfoque metaanalítico consolidado, realizada em sete passos. Antes de se seguir com os sete passos, foram determinadas as quatro principais áreas (*Business, Management, Economics e Business Finance*) com que o risco de crédito usando de *Big Data* tem maior interface, como forma de refinar a pesquisa.

A coleta de dados permitiu afirmar que a produção acadêmica do assunto é bastante recente, mas que o interesse acadêmico é crescente. A análise de autores e artigos permitiu visualizar que a produção de artigos é muito dispersa entre autores e países, porém, vale destacar, que a maior parte dos artigos foi publicada na China e República Tcheca. Os autores com maiores citações publicaram pouco, sendo a maioria dos trabalhos escritos por mais de um autor e a produtividade tem sido crescente, principalmente quando se trata de artigos que revisam e propõem modelos de pontuação de crédito para facilitar o acesso de pessoas físicas e jurídicas usando soluções de *Big Data* e, por outro lado, aprimorar as decisões de instituições financeiras na hora de conceder o crédito.

Finalmente, as linhas de pesquisa estão bastante voltadas para modelos de pontuação de crédito para facilitar o acesso de pessoas físicas e jurídicas usando soluções de *Big Data* e, por outro lado, aprimorar as decisões de instituições financeiras na hora de conceder o crédito. Portanto, seguindo o enfoque metaanalítico, foi possível enxergar o panorama dos estudos em risco de crédito

usando *Big Data* pelo mundo nos últimos dez anos, em que muitas mudanças ocorreram em apenas uma década.

Há limitações neste estudo que tangem a quantidade de palavras chaves na busca, para uma maior especificidade e segurança na análise sobre o que estava sendo pesquisado. Houve busca em apenas uma base de dados, o que restringiu ainda mais a quantidade de publicações encontradas sobre o tema.

Sendo assim, como sugestões de trabalhos futuros, recomenda-se a busca em mais combinação de mais palavras chave na busca, já que quando se fala de pontuação de crédito usando *Big Data* os termos de pesquisa variam muito e pode-se encontrar uma gama maior de conteúdos relacionados a esse, entretanto com nomes diferentes; busca em variadas bases de dados confiáveis; e sugere-se para estudos futuros procurar por limitações em soluções de *Big Data*.

6. REFERÊNCIAS

CALAZANS, A. T. S.; MARIANO, A. M.; PALDES, R. A. Uma revisão sistemática da bibliografia sobre métricas funcionais de tamanho de software utilizando o enfoque meta-analítico. **Universitas: Gestão e TI**, v. 5, n. 2, p. 12, 2015.

CAQUETTE, J. B.; ALTMAN, E. J.; NARAYANAN, P. **Gestão do Risco de Crédito: O próximo grande desafio financeiro**. Rio de Janeiro: Qualitymark, 1999.

CLARIVATE ANALYTICS. Web Of Knowledge, 2019. Disponível em: <http://wokinfo.com/training_support/training/web-of-knowledge>. Acesso em: 2 Outubro 2019.

DAMODARAN, A. **Gestão estratégica do risco**. Porto Alegre: Bookman, 2009.

EDVINSSON, E.; MALONE, M. S. **Capital intelectual: descobrindo o valor real de sua empresa pela identificação de seus valores internos**. São Paulo: Makron Books, 1998.

FIGUEIREDO, R. P. Gestão de Riscos Operacionais em Instituições Financeiras – Uma abordagem qualitativa. In: _____ **Dissertação de Mestrado, Universidade da Amazônia**. Belém (PA): [s.n.], 2001.

GHERARDI, C.; GHIEMMETTI, S. Escoragem de Crédito: Metodologia que identifica Estatisticamente o Risco de Crédito. **Tecnologia do Crédito**, v. II, 1997.

GUJARATI, D. **Econometria básica**. Rio de Janeiro: Elsevier, 2006.

INSTITUTE FOR SCIENCE INFORMATION. DIRECTORY INDEXING OF INTERNATIONAL RESEARCH JOURNALS, 2014. Disponível em: <<http://isithomsonreuters.org/main/evaluation>>. Acesso em: 2017 Maio 2017.

MARIANO, A. M.; CRUZ, R. G.; GAITÁN, J. A. Meta análises como instrumento de pesquisa: Uma revisão sistemática da bibliografia aplicada ao estudo das alianças estratégicas internacionais. **Congresso Internacional de Administração-Inovação Colaborativa e Competitividade**, 2011.

NEOWAY. Big Data Analytics: análise precisa para concessão de crédito, 2019. Disponível em: <<https://www.neoway.com.br/concessao-de-credito/>>. Acesso em: 2 Outubro 2019.

NEOWAY. O que é Big Data e qual a importância de implementá-lo na empresa, 2019. Disponível em: <<https://www.neoway.com.br/o-que-e-big-data/>>. Acesso em: 2 Outubro 2019.

OLIVEIRA, G. S. et al. **Fintech serviços financeiros: uma abordagem de serviços 4.0**. São Paulo: Anais do XLIII Encontro da ANPAD, 2019.

ONAY, C.; OZTURK, E. A review of credit scoring research in the age of Big Data. **Journal of Financial Regulation and Compliance**, 2018.

PARKINSON, K. L.; OCHS, J. R. Using credit screening to manage credit risk. **Business Credit**, 1998. 23-27.

RUSLI, E. M. Bad Credit? Start Tweeting, 2013. Disponível em: <<https://www.wsj.com/articles/SB10001424127887324883604578396852612756398>>. Acesso em: 7 Outubro 2019.

SAUNDERS, A. **Medindo o Risco de Crédito**: Novas abordagens para value at risk e outros paradigmas. Rio de Janeiro: Qualitymark, 2000.

SILVA, J. P. **Gestão e análise do risco de crédito**. 3ª. ed. São Paulo: Atlas, 2000.

SOUSA, A. F.; CHAIA, A. J. Política de crédito: uma análise qualitativa dos processos das empresas. **Caderno de Pesquisas em Administração**, v. 07, nº 03, p. 13-25, 2000.

THOMAS, L. C. **A survey of credit and Behavioural Scoring; Forecasting financial risk of lending to consumers**. Edinburgh, U.K.: University of Edinburgh, 2000.

VICENTE, E. F. R. A estimativa do risco na constituição da PDD. In: _____ **Dissertação de Mestrado, USP**. São Paulo: [s.n.], 2001.

WEI, Y. et al. Credit Scoring with Social Network Data. **Marketing Science**, 2016. 234-258.

WHARTON UNIVERSITY OF PENNSYLVANIA. A pontuação de crédito recorre aos dados “sociais” e prioriza o que considera importante, 2013. Disponível em: <<https://www.knowledgeatwharton.com.br/article/a-pontuacao-de-credito-recorre-aos-dados-sociais-e-prioriza-o-que-considera-importante/>>. Acesso em: 7 Outubro 2019.

ZUINI, P. Pequenas Empresas & Grandes Negócios: 10 Startups de Big Data que você precisa conhecer, 2015. Disponível em: <<https://revistapegn.globo.com/Startups/noticia/2015/07/10-startups-de-big-data-que-voce-precisa-conhecer.html>>. Acesso em: 2 out. 2019.