



Instituto Superior de Gestão

A análise preditiva sob o aspecto da regulação

Márcio Ozal de Abreu Monteiro

Dissertação de Mestrado para Obtenção do
Grau de Mestre em Estratégia de
Investimento e Internacionalização.

Orientador:

Professor Doutor Rui Moreira de Carvalho

Lisboa

2018

Instituto Superior de Gestão

A análise preditiva sob o aspecto da regulação

Márcio Ozal de Abreu Monteiro

Dissertação de Mestrado para Obtenção do
Grau de Mestre em Estratégia de
Investimento e Internacionalização.

Orientador:

Professor Doutor Rui Moreira de Carvalho

Lisboa

2018

Resumo

Esta investigação analisa o impacto das inovações disruptivas no tratamento de dados, e na extração de dados de alto valor para que as organizações possam melhorar o desempenho.

As novas tecnologias da área de análise preditiva, que contém um conjunto de técnicas de captação, tratamento da informação e seu produto de valor, que auxilia na mitigação dos custos de comercialização e propicia atingir o público cliente certo.

A análise preditiva, permitiu às empresas e organizações, criarem ou aperfeiçoarem seus produtos, para um público cada vez mais informado e em um mundo tão globalizado, o que torna a concorrência, não só inevitável, como uma variável comum. Através do tratamento de dados, as organizações conseguem distinguir o seu potencial cliente e perceber o que cada um quer, como e quando. Inovar e ser criativo, são características indispensáveis, e os dados é o novo “óleo” deste século, uma comparação a era exploratória do petróleo.

O novo Regulamento Geral de Proteção de Dados (RGPD), que por um lado, protege os direitos dos cidadãos e garante o respeito individual, por outro, ameaça o pilar importante das organizações em utilizarem técnicas de tratamentos de dados.

Este regulamento introduz recomendações, baseadas em princípios como consentimento, transparência, conceito de dados pessoais, “pseudoanonimato” (palavra criada no âmbito deste regulamento), direito a esquecimento e tem uma abrangência Europeia, mesmo aquelas empresas que não tenham atividade comercial na comunidade, mas que recolham dados na União Europeia.

O *blockchain*, sendo uma de várias das tecnologias disruptivas na história recente (e base de algumas criptomoedas, como o Bitcoin) pode ser uma alternativa, e através do seu modo de funcionamento, atender as recomendações do RGPD, e permitir às organizações o tratamento de dados. No entanto é preciso adequar.

Palavras-Chave: Big Data; Análise Preditiva; Regulamento Geral de Proteção de Dados; Blockchain.

Abstract

This research examines the impact of disruptive innovations in data handling, and in extracting high-value data so that organizations can improve performance.

The new technologies in the field of predictive analysis, which contains a collection of techniques of capture, information processing and its value product, which helps in mitigating marketing costs and propitiates to reach the right customer audience.

Predictive analysis has allowed companies and organizations to create or perfect their products for an increasingly informed public and in such a globalized world, which makes competition not only inevitable, but also a common variable. Through data processing, organizations can distinguish their potential client and realize what each one wants, how and when. Innovating and being creative are indispensable characteristics, and data is the new "oil" of this century, a comparison of the exploratory oil era.

The new General Data Protection Regulation (RGPD), which protects the rights of citizens and guarantees individual respect, on the other, threatens the important pillar of organizations in using data processing techniques.

This regulation introduces recommendations, based on principles such as consent, transparency, concept of personal data, "pseudo-anonymity" (word created under this regulation), right of forgetting and has a European scope, even those companies that do not have commercial activity in the community, but to collect data in the European Union.

Blockchain, being one of several disruptive technologies in recent history (and the basis of some crypto-coins, such as Bitcoin) may be an alternative, and through its mode of operation, comply with RGPD recommendations, and allow organizations to data. However, you need to adjust.

Key words: Big Data; Predictive Analytics; General Data Protection Regulation; Blockchain.

Agradecimentos

Meu primeiro agradecimento vai aquela pessoa que me deu a vida, que cuidou e me ensinou a ser um bom Homem. Agradeço e devo tudo o que sou a minha muita amada Mãe.

Sou muito grato também a duas pessoas que foram muito importantes, ao Carlos Gonçalves por me incentivar e apoiar para que eu pudesse estar no mestrado. Ao Professor Doutor Rui Moreira de Carvalho, por todo apoio que tem me dado para a realização da tese, com sua maneira ímpar de incentivar e orientar-me. Foi para mim um grande privilégio tê-lo como professor e orientador, estarei sempre grato.

Agradeço a minha namorada, que tem me apoiado e proporcionado todo o carinho.

Também não poderia de deixar de agradecer a todos os professores, colegas de curso, meus irmãos, amigos e colegas de trabalho.

Abreviaturas e acrónimos

AMA	<i>American Marketing Association</i>
AP	Análise Preditiva
B2B	<i>Business to Business</i>
B2C	<i>Business to Consumer</i>
BI	<i>Business Intelligence</i>
C2C	<i>Consumer to Consumer</i>
C&T	Ciência e Tecnologia
CdE	Conselho da Europa
CEO	<i>Chief Executive Officer</i>
CMO	<i>Chief Marketing Officer</i>
CNPD	Comissão Nacional de Proteção de Dados
DLT	<i>Distributed Ledger Technology</i>
IA	Inteligência Artificial
KDD	<i>Knowledge Discovery in Databases</i>
IBM	<i>International Business Machines</i>
P2P	<i>Peer-to-peer</i>
RGPD	Regulamento Geral sobre a Proteção de Dados
SI	Sistemas de Informação
TI	Tecnologia da Informação
UE	União Europeia

Glossário

Cloud - Conceito de “computação em nuvem”, refere-se à utilização da memória e da capacidade de armazenamento e cálculo de computadores e servidores compartilhados e interligados por meio da *Internet*.

Cookie - Arquivo texto que fica armazenado na máquina/computador do usuário todas as vezes que ele visita um site.

Big Data – Plataforma de tecnologias com capacidades de armazenamento e processamento em tempo hábil,

Digital Analytics - A análise dos dados qualitativos e quantitativos de ações e transações *online* feitas por uma empresa com ou sem comparação com os seus concorrentes, para direcionar uma melhoria contínua da experiência *online* dos atuais, e potenciais consumidores, a fim de produzir os resultados desejados pela empresa.

E-business – Envolve todas as ações *online* que impactam nos negócios da empresa.

E-commerce - Modelo de comércio que utiliza como base plataformas eletrônicas, como computadores, *smartphones*, *tablets* e etc.

Exabytes -

Google Analytics – O sistema gratuito de monitoramento de tráfego que pode ser instalado em qualquer *site*, loja virtual ou *blog*.

Petabytes -

Terabytes -

Web Analytics - Processo de medição, reunião, análise e a produção de relatórios de dados de navegação e interação com o objetivo de entender e otimizar o uso dos *sites* na *Internet*.

Webtrends - Ferramenta de *Web Analytics* que reúne e apresenta informações sobre o comportamento dos usuários em um *website*, *extranet*, *intranet*.

Índice

1. Introdução	1
1.1 Estrutura do Trabalho	4
2. Revisão da Literatura	5
2.1 <i>Big Data</i>	5
2.2 Análise Preditiva.....	8
2.3 Regulamento Geral de Proteção de Dados	12
2.4 <i>Blockchain</i>	16
3. Enquadramento da análise preditiva como ferramenta de apoio à decisão	21
3.1 A análise preditiva como ferramenta de apoio à tomada de decisões	24
3.2 Casos de aplicação de análise preditiva.....	27
4. Modelo de Investigação	30
4.1 Recolha de Informação	30
4.2 Perguntas de partida.....	30
4.3 Instrumentos	32
5. A Análise Preditiva após a implantação do Regulamento Geral de Proteção de Dados 34	
5.1 <i>Blockchain</i> e sua multifuncionalidade.....	35
5.2 <i>Blockchain</i> , análise preditiva e o Regulamento Geral de Proteção de Dados	38
6. Conclusões	43
6.1 Limitações e sugestões para investigação	45
7. Bibliografia	46

Índice de Quadros

Tabela 1 - Características do Big Data	7
--	---

Índice de Figuras

Figura 1 - Linha do tempo da evolução dos SI nas empresas.....	7
Figura 2 - Estimativa de pessoas com acesso à internet	10
Figura 3 - Volume de informação por tipo de serviço da Internet em 2017.....	11
Figura 4 - Exemplo das Criptomoedas que se utilizam da <i>Blockchain</i> , e a independência entra as camadas.	17
Figura 5 - Exemplo de uma DLT e a comunicação P2P.	18
Figura 6 - Nível de interesse por <i>Big Data</i> em pesquisas pelo item. Medido mediante pesquisas na página da Google.	25
Figura 7 – Hierarquia de aprendizado.	26
Figura 8 - Google Trends - Nível de interesse por <i>Blockchain</i> em pesquisas pelo item. Medido mediante pesquisas na página da Google.	35
Figura 9 – Infográfico: Diversas utilizações para o <i>Blockchain</i>	37

1. Introdução

O direito a privacidade, preservação do direito a dignidade e de escolha tem sido debatido pela sociedade, principalmente após o surgimento e crescimento das redes sociais. O Parlamento Europeu acompanhou a necessidade de uma regulamentação para atender os interesses comuns dos indivíduos ao estabelecer o Regulamento Geral de Proteção de Dados (RGPD) 2016/679 do Parlamento Europeu.

Esta investigação tem como objetivo estudar a influência da regulamentação da União Europeia (UE) 2016/679 do Parlamento Europeu (que revogou a Diretiva 95/46/CE) sobre o tratamento de dados pessoais e os efeitos introduzidos pelo regulamento à análise preditiva (AP).

O regulamento invoca a necessidade de adaptação por parte das empresas, e demais organizações, ao processo de utilização dos dados nos processos de tomadas de decisão.

Assim, importa começar por uma melhor definição dos conceitos. Dado, é definido por Miranda (1999, p. 285) como um conjunto de registos qualitativos ou quantitativos conhecidos, organizados, agrupados, categorizados e padronizados adequadamente, que tende a transformar-se em informação.

A AP tem ganhado terreno tornando-se uma peça fundamental do movimento de *Big Data* e *Internet of Things*. No contexto atual, existe uma quantidade de informação “deixada” pelas pessoas que permitem a empresas como a Facebook, tratar estes dados através de algoritmos informáticos e detetar padrões, o que permite, por exemplo, campanhas de marketing direcionadas a um público “escolhido”, o que pode ser mais atrativo e recompensador para as empresas.

Em 2012, a Forbes apresentou o repto “*Is Data The New Oil?*” (Rotella, 2012). A intenção era atrair atenção às empresas para uma espécie de “novo combustível” para a competitividade: Tratar os dados como um ativo corporativo, permitindo o compartilhamento de dados relativo a clientes e produtos e serviços. Desta forma, o “compartilhamento” de dados de produtos pode abrir oportunidades para criar produtos novos e inovadores, otimizar processos de fabrico, nicho de mercado e etc.

Já em 2018, a Forbes publicou uma reflexão ao texto anterior: “*Here's Why Data Is Not The New Oil*” (Marr, 2018) Em que sugeria que “os dados também se tornam mais úteis

quando reaproveitados, sendo que uma vez processados, os dados geralmente revelam novos *insights*, descobrir padrões ou mesmo situações que poderiam escapar a vista de um “olho humano”, é algo que os algoritmos de AP buscam colmatar, maximizando assim o valor da informação”.

As técnicas de AP (como aprendizado de máquina, algoritmos de inteligência artificial e etc) permitem às empresas uma otimização de processos, custos, melhoramentos em produtos e outras vantagens. No entanto devem enquadrar a atividade de tratamento de dados para que estejam coerentes com as regras dos direitos de defesa e liberdade fundamentais das pessoas impostas pelo RGPD.

De acordo com Quivy e Campenhoudt (2011), o trabalho de investigação é, por definição, algo “que se procura”, caminhando para um conhecimento cada vez mais elevado. Por esta razão, deve-se estabelecer um fio condutor tão claro e coerente quanto possível (Torres, 2016).

Yin (1994) refere que “a definição das questões a investigar é o passo mais importante num estudo de investigação”. Parlett e Hamilton (1976), por outro lado, defendem a “focalização progressiva” como base fundamental para o processo evolutivo da clarificação do problema (Carvalho, 2008).

Nesse sentido, foram sugeridas as seguintes perguntas de partida:

Q1 *A análise preditiva é relevante para as organizações?*

Considerando as tecnologias emergentes de *Big Data*, mas especificamente, da evolução da AP, é importante correlacionar a aplicabilidade e o possível aporte de valor às organizações/empresas.

“Com sucessivas invenções, como as do telégrafo, do telefone, da rádio, da televisão e dos computadores, a quantidade de dados no mundo cresceu rapidamente ao longo do século XX” (Ross, 2016, p. 168). Outros eventos como o surgimento das redes sociais, embalados pelo combustível da globalização, proporcionam um grande volume de informação nos dias atuais que são gerados e armazenados. “Noventa por cento dos dados digitais do mundo inteiro foram gerados nos últimos dois anos” (Ross, 2016, p. 169).

Possibilitar às organizações se tornarem mais competitivas e inovadoras, passa muito pelo processo de tomada de decisões onde a capacidade de processamento e tratamento de dados é importante.

Importa, acima de tudo, que estas questões de investigação sejam suficientemente claras, realistas e pertinentes (Quivy e Campenhoudt, 2013, p. 34-43)

Q2 *Qual a importância da RGPD nas organizações?*

A adoção e a utilização das redes sociais fazem com que as pessoas deixem traços de interesses, personalidade, relações, etc. Assim, a necessidade da regulação aparece como interesse comum para a defesa dos direitos, no fundo, para a dignidade das pessoas. A necessidade de transparência do processo de tratamento de dados relativo a pessoas é importante para a garantia dos direitos e pode ser um elemento importante para o aumento da confiança na empresas e organizações.

Alguns recentes eventos como a eleição do último presidente americano, e a decisão do *Brexit*, popularizaram o tema da privacidade e aumento ou debate na sociedade.

Segundo o texto de informação da Comissão Europeia, a regulamentação “tornará os negócios mais fáceis para as empresas em toda a UE. A reforma ajudará as empresas a aproveitarem plenamente a economia digital em todo o mercado único digital da UE” (Europeia, 2018b). A transparência e o estado de *compliance* à regulamentação, não é apenas obrigatório, mas também é instrumento para elevar a confiança, essencial no estabelecimento da relação com os compradores de bens, produtos e serviços.

A regulamentação é, talvez, uma oportunidade, mas que apresentam desafios, sejam a adaptar os sistemas informáticos, e da garantia do tratamento de informações. Com o objetivo de permitir a continuidade no tratamento de dados.

Q3 *A tecnologia blockchain, pode assegurar a realização de tratamento de dados face ao RGPD?*

“A realidade globalizante da economia europeia, em todos os aspetos e também na questão dos dados pessoais, aliada à descoberta, já não tão recente, de que os dados pessoais possuem um valor económico intrínseco, têm levado as organizações a fazerem grandes investimentos na recolha e tratamento de dados de milhões de consumidores europeus” (Saldanha, 2018).

Conquistar a confiança de um público exigente e com acesso rápido a informação é uma tarefa que se deve ter alguma cautela. Garantir a segurança dos dados e estar de acordo com regulamentação é preocupação das organizações. Empresas e organizações que não estejam *compliance* com a regulamentação, necessitam investir tempo e dinheiro. Não sendo os recursos ilimitados, é preciso prudência.

“Apesar de muitos elementos secundários associados à *bitcoins* terem sido pirateados, a tecnologia *blockchain* em si nunca foi comprometida” (Ross, 2016, p. 116). A segurança da informação é algo importante não só pela questão em si, mas também pelas regras introduzidas pela regulamentação. “A tecnologia *blockchain* fornece precisamente infraestruturas que permitem ter essa confiança” (Martins, 2018).

1.1 Estrutura do Trabalho

A investigação encontra-se estruturada em seis capítulos. O primeiro trata a introdução, e o modelo conceptual adotado para o desenvolvimento deste trabalho.

No segundo capítulo é feita a revisão da literatura que foca a introdução dos conceitos mais relevantes para a investigação, nomeadamente a (i) *Big Data*, (ii) Análise Preditiva, (iii) Regulamento Geral de Proteção dos Dados e (iv) *Blockchain*.

No terceiro capítulo é feito o enquadramento da análise preditiva e da sua dinâmica no enriquecimento do processo de tomada de decisão.

O Modelo de Investigação é apresentado no capítulo quarto.

O capítulo quinto trata da análise do objeto de estudo, a relação da análise preditiva no âmbito do RGPD e como o *blockchain* pode auxiliar na resposta ao regulamento

Finalmente, são aduzidas as conclusões (capítulo seis), expostas as limitações e por último são dadas pistas para futuros projetos de investigação.

2. Revisão da Literatura

Com o objetivo de estudar a temática do processo de AP em um aspeto geral, e sob o a influência da regulamentação europeia.

Neste sentido, vamos dar particular ênfase ao estudo dos conceitos de (i) *Big Data*, (ii) Análise Preditiva, (iii) Regulamento Geral de Proteção dos Dados e (iv) *Blockchain*.

2.1 *Big Data*

Big Data, surge como um conjunto de tecnologias em que o processo de se associarem permite captar e processar grandes volumes de dados em tempo “próximo” do real possibilita extrair informação de valor acrescido (foco no tratamento dos dados). Os consumidores são exigentes e possuem acesso a muita informação. “Com a globalização, e com a diretiva de que nada deve ser produzido, transportado ou comprado antes da hora exata chamado *Just in time*, a expansão virtual se tornou necessária” (Tessarolo & Magalhães, 2014).

Entrevistas feitas durante o evento Marketing2020¹, permitiu a Lindstrom a sumarizar a seguinte reflexão: “Millward Brown Vermeer iniciou recentemente o Marketing2020, um dos estudos de liderança de marketing mais abrangentes alguma vez lançados, e que inclui entrevistas aprofundadas com mais de 350 CEOs, CMOs e diretores de agências. Não foi, portanto, surpreendente que os autores Marc de Swaan Arons, Frank van den Driest e Keith Weed tenham descoberto que muitas organizações de marketing se perderam do seu caminho. Num artigo publicado na *Harvard Business Review*, os autores concluíram que se os dados e análise se incluem na categoria pensar e o conteúdo, *design* e desenvolvimento de produção se incluem na categoria fazer, então os especialistas de marketing que se concentram no envolvimento e integração dos dados *online e offline* – ou seja, o casamento dos megadados com os pequenos dados – é um ingrediente crucial para a sobrevivência do marketing e para o sucesso do século XXI.” (Lindstrom, 2017).

¹ Iniciativa de liderança em marketing global, com o foco no alinhamento de estratégias, estrutura e capacidades para proporcionar o crescimento de negócios. Possui um conjunto de boas práticas. Acedido em: 23 de abril de 2018, em: <http://www.mbvermeer.com/portfolio-itemmarketing2020/>.

O interesse por tecnologias da informação nas questões de armazenamento e as potencialidades que o tratamento de dados e AP podem acrescentar valor as empresas e organizações foi o tema central da pesquisa realizada no Marketin2020.

O termo *Big Data*, foi introduzido por um analista da *Gartner*², em meados dos anos 2000. O termo passou a ser comum em nosso cotidiano, “de um simples neologismo para um conceito capaz de definir o futuro de empresas e organizações no que diz respeito à análise e estruturação de dados.” (Volpato, Rufino, & Dias, 2013).

O termo *Big Data* refere-se a um conjunto de tecnologias, formando um ecossistema e possui algumas características, nomeadamente:

- a) “*Big Data*, em geral, é definido como ativos de alto volume, velocidade e variedade de informação que exigem custo-benefício, de formas inovadoras de processamento de informações para maior visibilidade e tomada de decisão”. (Pemberton, 2018).
- b) “...as tecnologias de *Big Data* descrevem uma nova geração de tecnologias e arquiteturas projetadas para extrair economicamente o valor de volumes muito grandes e de uma variedade de dados, permitindo alta velocidade de captura, descoberta e/ou análise”. (IDC, 2011)
- c) “*Big Data* = volume + variedade + velocidade + veracidade + valor”. (Taurion, 2013).
- d) “*Big Data* é o termo utilizado para descrever grandes volumes de dados e que ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas a cada dia”. (IBM, 2013).

Big Data apresenta-se como uma plataforma tecnológica com capacidades de armazenar grandes volumes de dados (textos, fotos, vídeos, livros e etc) viabilizando o tratamento de dados com grande velocidade. Observa-se que estas definições recordam as questões centrais e abordadas por Taurion (2013), quando refere como *Big Data* as tecnologias que atendam 5 variáveis/características, conhecidos vulgarmente como os 5Vs.

² Importante empresa de consultoria que desenvolve tecnologias que auxiliam nas tomadas de decisões.

A análise preditiva sob o aspecto da regulação.

Podemos avaliar estas variáveis ou características que definem das tecnologias *Big Data* através do quadro apresentado de seguida:

Tabela 1 - Características do *Big Data*

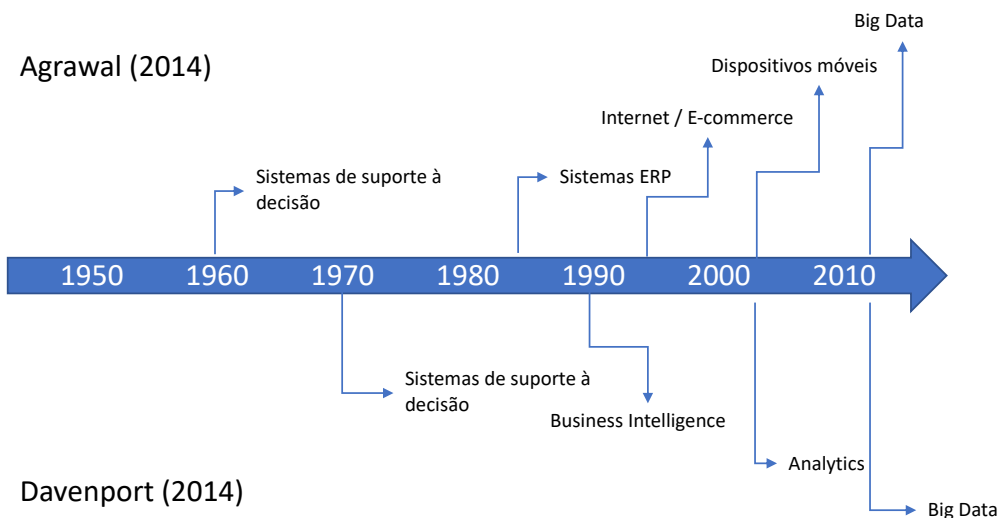
Características / Descrição
Volume: (i) Relacionado ao tamanho e quantidade de dados; (ii) Centenas de Terabytes ou Petabytes; (iii) Grande quantidade e complexidade de dados; (iv) 2,5 Exabytes de dados criados por dia, dobrados a cada 40 meses;
Velocidade: (i) Necessidade de respostas em um curto prazo ou em tempo real; (ii) Dinâmica de crescimento e processamento de dados; (iii) Dados capturados e processados quase em real time; (iv) Velocidade de captura e análise de dados, formando um fluxo contínuo;
Variedade: (i) Diversidade de origens, formas e formatos de dados; (ii) Grande variedade de fontes e formas de dados com o desafio de encontrar os padrões de dados úteis para os negócios;
Veracidade: (i) Autenticidade, reputação de origem e confiabilidade dos dados;
Valor: (i) Dados com significado para os negócios, que contribuam com valor agregado;

Fonte: Adaptado de “Uso Corporativo do Big Data: Uma revisão de literatura” (Silveira, Marcolin, & Freitas, 2015).

Da leitura é possível inferir que estas características são de (i) grande interesse por parte das organizações; (ii) possuem a necessidade de informatização, (iii) permitem atender as necessidades de negócio e (iv) auxiliam a otimização de recursos.

Assim, para Agrawal (2014) e Davenport (2014), o *Big Data* assume-se como uma evolução tecnológica nas organizações, como podemos ver pela cronologia abaixo (

Figura 1 - Linha do tempo da evolução dos SI nas empresas.



Fonte: Adaptado de “Uso Corporativo do Big Data: Uma revisão de literatura” (Silveira *et al.*, 2015).

O quadro cronológico apresentado pelos autores permite-nos visualizar diversas fases evolutivas das plataformas digitais de apoio a decisão. Apesar de alguma divergência cronológica entre Agrawal (2014) e Davenport (2014), todas as áreas de conhecimento e eventos apontados foram importantes para a consolidação das plataformas *Big Data*.

São técnicas informáticas de processamento de dados (suportados em modelos matemáticos e estatísticos) e que hoje convergem na plataforma *big data*, desde os sistemas de ERP (*Enterprise Resource Planning*) que tinham como objetivo conter todas as funções de uma empresa, desde a contabilidade, recursos humanos, stocks, etc.

Na realidade, os sistemas de ERP foram as primeiras plataformas a terem um volume de informação e com potencial de criar valor para aplicações de *Business Intelligence* (que se baseia em modelos estatísticos e analíticos para a normalização e tratamento dos dados), que auxiliavam no processo de análise e apoio a decisão.

Outros dois pontos marcantes, foram a internet e o *e-commerce* (dado ao início do comércio eletrónico), nos anos 2000, que acabaram por estar na génese da crise conhecida como ponto com (.com), mas que popularizaram a internet e a venda de produtos e serviços de forma *online*. O aparecimento dos dispositivos móveis, a partir do *smartphone*, popularizou o acesso fácil e prático a conteúdos da Internet e agilizou comunicação.

Empresas como a *Google, Facebook, Twitter* e outras, contribuíram para a popularização massiva das redes sociais através de algoritmos e tecnologias de armazenamento e tratamento de dados, permitindo às empresas a comercialização do acesso a “dados” para que possam desenvolver ações de marketing ou outras com fins comerciais, ou outros (sócio-políticos, etc).

2.2 Análise Preditiva

O termo preditivo remete a capacidade de antecipar. A análise preditiva possibilita avaliar um conjunto de dados, detetar padrões e apontar alguma informação importante que permita a empresa traduzir “tendências” em resultados quantitativos.

Com o crescimento no volume de dados gerados e a capacidade de armazenamento, é interessante para as organizações desenvolverem técnicas de processamento de dados que

permitam extrair valor, seja na otimização de processos, seja na aproximação com o cliente para determinar o comportamento de consumo e de mercado. Há muitas áreas que a informação tratada pode ter impactos significativos (positivos, ou, outros).

Este cenário remete-nos a uma analogia importante para entendimento de algumas técnicas informáticas para o processamento de dados, sendo que a capacidade de armazenar grandes volumes de dados pode promover a analogia comum a “mina de ouro”, onde é necessário a realizar processos de pesquisa e mineração que possibilite extrair “aquilo que é mais valioso”.

As técnicas de *data mining*³ possibilitam tratar um conjunto de dados e extrair informações que agreguem valor ao negócio e/ou processo sendo que “os maiores benefícios pela execução de processos de mineração de dados⁴ é a criação de inteligência de negócios sobre determinado assunto, também referenciado como KDD - *Knowledge Discovery in Databases*⁵ – ou descoberta de conhecimento sob bases de dados” (Gonçalves, Pereira, Pinheiro, & Aguiar, 2008)

Estas técnicas e tecnologias são desenvolvidas com objetivo de suportar os processos de tomadas de decisão promovendo potenciais ganhos de competitividade nas empresas. Desta forma assumem-se como ferramentas importantes num cenário de crescente globalização e conseqüente interligação de pessoas e bens.

Para (Novo & Neves, 2013), “a inteligência analítica está presente em empresas dos mais diversos segmentos, seja no cerne do negócio de multinacionais como *Google* ou *Walmart*, seja oferecendo *insights* para áreas específicas como marketing e finanças de empresas de diversos segmentos. A inteligência analítica permite melhorar o poder de competição das empresas, além de ser uma parte importante dos sistemas de *business intelligence*⁶ (BI)”

³ “É o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detetar relacionamentos sistemáticos entre variáveis, detetando assim novos subconjuntos de dados.” Wikipédia, acessado em 28/04/2018, em: https://pt.wikipedia.org/wiki/Minera%C3%A7%C3%A3o_de_dados.

⁴ É o mesmo que *Data Mining*. Gonçalves (2001, p. 10) afirma: “pode ser considerada o núcleo da descoberta de conhecimento em bases de dados, consistindo na aplicação de algoritmos de extração”.

⁵ FAYYAD (1996), define como o “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”.

⁶ “É um conjunto de técnicas e ferramentas para auxiliar na transformação de dados brutos em informações significativas e uteis a fim de analisar o negócio.” Wikipédia, acessado em 25/04/2018, em: https://pt.wikipedia.org/wiki/Intelig%C3%A2ncia_empresarial.

“As empresas estão nadando em um mar de dados, os quais são muito volumosos ou muito complexos para serem analisados da forma tradicional” (Franks & Davenport, 2012). Na mesma linha (Bughin, Livingston, & Marwaha, 2011) sugerem que “a análise dessa gigantesca massa de dados está, rapidamente, a evoluir de modo exponencial tornando-se a nova fronteira da competição empresarial”

As plataformas tecnológicas de *big data* surge com modelos evolutivos da análise de dados tradicionais, oferecendo maior capacidade para processamento de informações de apoio a decisão e que agregue valor.

Nesse sentido, a AP tem apoiado a capacidade de análise aos dados de diversas fontes, formatos e de maior volume permitindo que “empresas que utilizam a análise de dados como ferramenta podem beneficiar da melhoria de processos produtivos, de gestão, de operação ou em seus modelos de vendas e interação com os consumidores” (Novo & Neves, 2013).

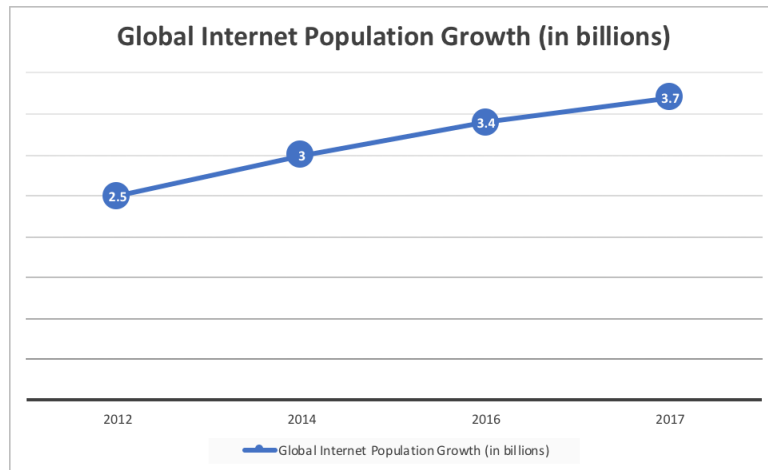
Freitas (2000) sugere que a informação “serve” à tomada de decisão; logo, a necessidade de decidir com maior precisão é justificada pela necessidade que se tem de agir dentro das organizações e no campo da pesquisa. A aplicação de técnicas e algoritmos que permitam realizar uma análise a grandes volumes de dados em uma plataforma de *big data*, permite minimizar os riscos resultantes de uma tomada de decisão inadequada, logo, é fundamental que os gestores possam embasar com informações seguras, relevantes e que possam antecipar situações.

O processo de engajamento à internet vem a crescer ao decorrer dos anos (como se vê na figura seguinte), sendo que o volume de dados gerado a “todo minuto” aumenta, pelo que Siegel (2016) sugere que “deixamos pistas da nossa identidade à vista de todos, mas elas são universais e, numa era digital, também são indelévels.” Estes rastros possui um potencial de informação e quanto maior o nível de adesão ou universo de dados, maior a taxa de exatidão de detecção de padrões e avaliações estatísticas.

Em divulgação recente, a DOMO, aponta que a internet está ao alcançar uma grande fatia populacional do nosso planeta, como se pode ver no gráfico evolutivo abaixo:

Figura 2 - Estimativa de pessoas com acesso à internet

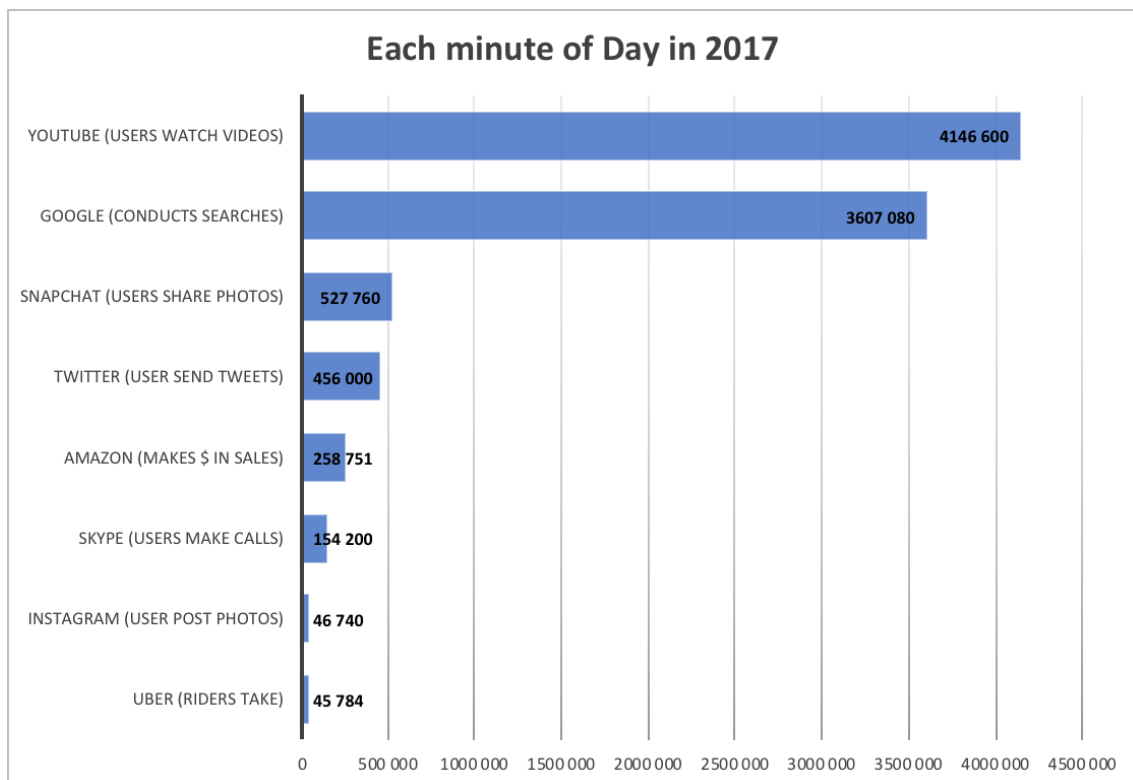
A análise preditiva sob o aspecto da regulação.



Fonte: DOMO, 2017. <https://www.domo.com/learn/data-never-sleeps-5>

A figura da sugere que as pessoas estão a interagir mais com Internet (seja a criar conteúdos digitais, realizar compras *online* e etc), sendo possível quantificar o volume de informação, dos diversos formatos, e podem indicar as rotinas, características e personalidade de cada pessoa, entre outros fatores mensuráveis.

Figura 3 - Volume de informação por tipo de serviço da Internet em 2017



Fonte: DOMO, 2017. <https://www.domo.com/learn/data-never-sleeps-5>

O trabalho da DOMO (2017) ilustra o significativo aumento no número de utilizadores (figura 2), e do volume de informação por tipo de plataforma digital (figura 3). Assim, as variadas plataformas digitais (redes sociais, *streaming* e etc) oferecem, em média a cada

minuto cerca de 4 milhões de vídeos são visualizados no *Youtube*, realizam cerca 3,6 milhões de pesquisas no motor de busca da *Google*, e publicam aproximadamente 600 novos artigos na plataforma *Wikipédia*.

Desta forma, a adoção de uma plataforma *Big Data* aparenta atender as necessidades de tratamento de dados, sendo possível aplicar técnicas de processamentos e mineração de dados para a produzir “informações de valor” potenciando oportunidades, ou otimização, de negócios.

À guisa de conclusão, é possível afirmar que análise preditiva pode ser uma ferramenta importante para o cumprimento deste objetivo, pois mitiga falhas ou não perceção de padrões em um volume maior e mais diversificado, a partir de algoritmos e modelos matemáticos.

2.3 Regulamento Geral de Proteção de Dados

O regulamento surge como demanda dos cidadãos como defesa da privacidade. As redes sociais revolucionaram em muito esta matéria, pois contêm os rastros e traços das personalidades e interesses pessoais através da quantidade de dados relativo a pessoas disponíveis em meio digital. Alguns recentes acontecimentos como as últimas eleições americanas (Donald Trump eleito presidente dos Estados Unidos em 2016) e as eleições do *Brexit*⁷, no Reino Unido, ilustram como uma empresa, neste caso a *Cambridge Analytica*⁸, conseguiu processar, extrair e utilizar a informação em proveito de terceiros.

“A investigação do jornal britânico *The Guardian* descobriu que a empresa de análise de dados *Cambridge Analytica* recolheu informações pessoais de cerca de 50 milhões de pessoas através da rede social, para ajudar o republicano Donald Trump a vencer as eleições presidenciais de 2016 nos Estados Unidos, provocou um tobo das ações, que chegaram a negociar abaixo dos 150 dólares (cerca de 122,8 euros).” (Ferreira, 2018) para o *Jornal Económico*⁹

Estes acontecimentos ilustram o potencial que estas técnicas podem produzir. Como

⁷ Alcinha para o plebiscito que votou pela saída da Grã-Bretanha do tratado da UE. É referido popularmente como a junção da palavras *Britain* e *Exit*.

⁸ *Cambridge Analytica* é uma empresa privada que combina mineração e análise de dados com comunicação estratégica.

⁹ Acedido em 26/04/2018 em: <http://www.jornaleconomico.sapo.pt/noticias/facebook-enfrenta-novo-teste-lucros-resistem-ao-escandalo-cambridge-analytica-297700>.

indica Machado¹⁰ (2018), em uma nota de um jornal: “O Brexit nunca teria acontecido sem a *Cambridge Analytica*”, em que completa a informação como sendo a de um ex-funcionário (Machado, 2018).

“Através dos dados que recolheu de perfis de redes sociais, com maior enfoque no *Facebook*, criou publicidade específica para alterar ou tentar definir o sentido de voto dos utilizadores” (Machado, 2018). Está na génese das redes sociais de ser uma plataforma de aproximação de pessoas, contactos, e partilha de informação entre amigos. Hoje conhecemos alguns dos perigos e da forte influência e capacidade de afetar um Estado e toda uma nação.

Com o objetivo de harmonizar interesses das empresas e organizações com o direito da sociedade (na proteção do indivíduo) foi promovido o Regulamento Geral de Proteção de dados (RGPD). Trata-se de uma lei da União Europeia, de carácter obrigatório e poder jurídico com o objetivo de proteger cidadãos europeus independente da geolocalização da empresa que detenha dados de indivíduos de cidadania europeia. Em Portugal, será regulado pela Comissão Nacional de Proteção de Dados (CNPD).

“Mais de 90% dos europeus afirmam pretender que sejam gozados os mesmos direitos de proteção de dados em todos os países da UE e 8 em cada 10 pessoas sentem que não têm o controle total sobre a utilização que é feita dos seus dados pessoais”, aponta o Eurobarómetro¹¹ em 2015 e citado em uma comunicação oficial da União Europeia (Europeia, 2018b).

Portanto, a RGPD afeta os 28 Estados-membros da UE e a empresas/organizações de qualquer país que venda produtos ou serviços dentro da UE ou que simplesmente possuam dados de cidadãos europeu. A Comissão Europeia fundamentou o RGPD em princípios que certifiquem que os direitos dos cidadãos se mantenham efetivos no contexto digital, e que permita construir uma relação de confiança com as empresas por meio da transparência na utilização de dados.

A nova regulamentação, descreve os dados pessoais como: “informações relativas a um

¹⁰ Jornal Observador, acessado em 28/03/2018, em: <https://observador.pt/2018/03/26/o-brexit-nunca-teria-acontecido-sem-a-cambridge-analytica-revela-ex-funcionario/amp/>.

¹¹ São estudos realizados nos Estados-membros, para melhorar a percepção e expectativas dos cidadãos europeus. Para mais informações, aceder em: http://www.europarl.europa.eu/portugal/pt/informacao_ue/eurobarometro.html.

indivíduo vivo identificado ou identificável. Por exemplo: nome, apelido, endereço, endereço de correio eletrónico, dados de localização. (interpretação do artigo 4º, nº 1 do regulamento)” (Europeia, 2018a).

Dentro dos direitos garantidos pelo regulamento e conforme publicação da UE em nota¹², é possível destacar:

- **Transparência:** Direito de receber informações claras e compreensíveis sobre quem efetua o tratamento de dados do indivíduo, quais dados estão a ser utilizados no processamento/tratamento e quais os fins deste tratamento, conforme indicado nos artigos 12º-14º do regulamento.
- **Acesso facilitado aos dados pessoais:** O direito de pedir o acesso aos dados pessoais detidos por uma organização, refere o artigo 15º do regulamento.
- **Transmissibilidade:** Garante o direito ao cidadão, exclusiva, de permitir a um prestador de serviços que transmita os seus dados pessoais a outro prestador de serviços, conforme o artigo 20º do regulamento.
- **Esquecimento:** Garante ao cidadão o direito de ser esquecido, ou seja, poderá solicitar que os seus dados pessoais sejam apagados caso pretenda cessar o seu tratamento e se não houver motivos que justifiquem a sua conservação por parte da empresa que os detém. Por exemplo, se, ao inserir o seu nome num motor de pesquisa em linha, verificar que os resultados incluem ligações para um artigo de jornal antigo relativo a uma dívida que já pagou há muito tempo, poderá pedir ao motor de busca que apague essas ligações (esta medida só é aplicável caso não seja uma figura pública ou caso o seu interesse na supressão do artigo prevaleça sobre o interesse do público em geral de ter acesso à informação). Artigo 17º do regulamento.
- **Segurança dos dados por defeito:** Nos casos em que as empresas necessitam do consentimento do cidadão para tratarem os seus dados, terão de o solicitar e de indicar claramente qual a utilização que será dada aos seus dados pessoais. O seu consentimento tem de refletir os seus desejos de forma inequívoca e deve ser dado em forma de uma ação afirmativa. Deste modo, as empresas não poderão

¹² Reforma em matéria de proteção de dados na EU: *proteção de dados reforçada para os cidadãos europeus*. Janeiro de 2018, acedido em 24/04/2018 em: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_pt.

esconder-se atrás de extensos termos e condições legalistas que nunca chegou a ler. Fundamentado nos artigos do regulamento: artigo 4º (nº 11) e artigo 7º.

- **Comunicação em casos de quebras de segurança:** Se os seus dados forem perdidos ou roubados, e se esta violação de dados for suscetível de constituir um prejuízo para si, a empresa que esteve na origem da violação de dados terá de o informar (e de informar a autoridade de controlo competente em matéria de proteção de dados) sem demora injustificada. Se a empresa não o fizer, poderá ser-lhe aplicada uma coima. Conforme texto dos artigos 33º-34º do regulamento.
- **Proteção aos dados de menores:** As crianças podem estar menos cientes dos riscos e das consequências da partilha de dados e tem menos conhecimento dos seus direitos. É por esta razão que qualquer informação dirigida especificamente a uma criança deve ser adaptada para ser facilmente acessível e formulada numa linguagem clara e simples. Indicado no artigo 8º do regulamento.

O regulamento obriga as organizações a estarem em conformidade, e de forma explícita, a repensarem sobre os modelos adotados para o tratamento de dados através da adaptação dos sistemas informáticos e dos meios de comunicação é crucial. Por outro lado, oferece a oportunidade às empresas de procurarem o ganho de competitividade através da utilização dos dados, convertidos e informação, como “minas de ouro”, de forma responsável.

Este regulamento surge, pois, em substituição a antiga Diretiva de Proteção de Dados. “Esta regulação cria novos direitos individuais e impõe novas medidas de responsabilização sobre organizações que coletam ou processam dados” (Rubinstein, 2013), que continua a afirmar que, “... o *tsunami* Big Data provavelmente sobrecarregará esses esforços de reforma.

Assim, uma abordagem suplementar deve ser considerada na perspetiva da utilização de códigos de conduta ou de ética. Na realidade, os reguladores devem encorajar as empresas a adotar novos modelos de negócios com base no empoderamento do consumidor, oferecendo incentivos como flexibilidade regulatória e penalidades reduzidas.”

O RGPD que entrou em vigor a 25 de maio de 2018.

A Gartner¹³, divulgou, em 2018, uma lista de 6 tecnologias, que sugere como essenciais para a competitividade das organizações. E para AP, sugeriu que: “a análise preditiva é uma variante do aprendizado de máquina que antecipa comportamentos futuros e estima resultados desconhecidos. Os líderes de marketing exploram seu potencial de gerenciamento de rotatividade, vendas cruzadas, propensão à compra, gerenciamento de campanhas multicanal e previsão do valor da vida útil do cliente, bem como novos aplicativos para melhorar a tomada de decisões nos negócios.” (Pemberton, 2018).

2.4 Blockchain

“Blockchain teve sua origem no ano de 2008, juntamente com a proposta da criptomoeda¹⁴ *Bitcoin*¹⁵. Esse conceito tecnológico tornou-se a fundamentação de praticamente todas as criptomoedas que surgiram desde então.

Apesar de sua concepção original para pagamentos eletrônicos, a tecnologia *Blockchain* é de aplicabilidade bem mais abrangente” (Rodrigues, 2017).

“A *blockchain* é uma base de dados distribuída, que é usada para guardar uma lista crescente de registros, chamados de blocos, e que são encadeados numa cadeia” (Ciccarino & Araki, 2017).

Devido ao seu modelo de funcionamento, a aplicação desta tecnologia para outras finalidades parece ser viável. Assim, a *Blockchain* “é uma plataforma em forma de livro-razão que surgiu com o intuito de resolver o problema do duplo-gasto (em inglês, *double spending problem*¹⁶)” (Ulrich, 2014).

A imagem seguinte demonstra as plataformas digitais necessárias para viabilizar o funcionamento das criptomoedas, tais como *Bitcoin* e *Ethereum*. São necessárias a associação de tecnologias que garantam a confiança, protocolos que definem a forma de

¹³ Gartner é uma empresa de consultoria, e que desenvolve tecnologias relacionadas a introspeção necessária para seus clientes tomarem suas decisões todos os dias.

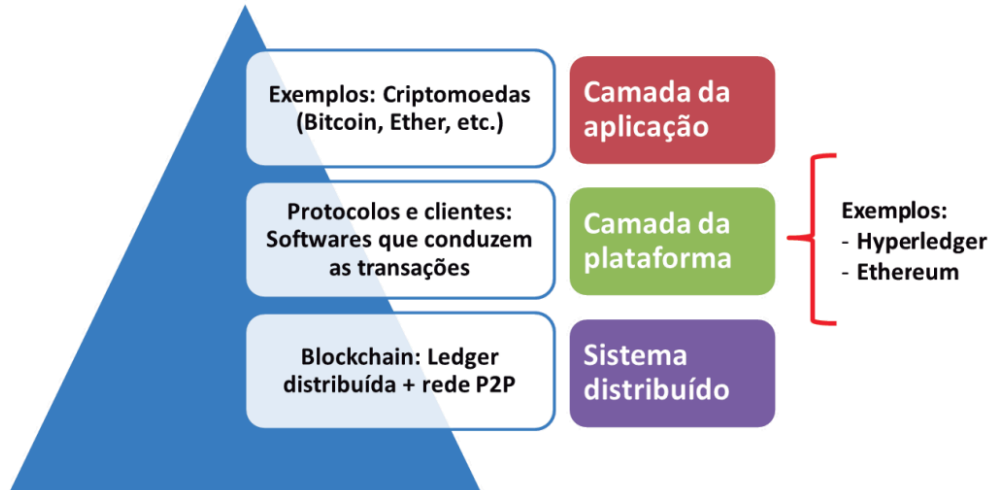
¹⁴ São moedas digitais, descentralizadas e que se utiliza da *blockchain* e criptografia para assegurar a validade das transações.

¹⁵ “Bitcoin é uma moeda digital do tipo criptomoeda descentralizada e, também um sistema econômico alternativo, apresentada em 2008 na lista de discussão *The Cryptography Mailing* por um programador, ou um grupo, de pseudônimo Satoshi Nakamoto.”, acessado 30/04/2018 em <https://pt.wikipedia.org/wiki/Bitcoin>.

¹⁶ O gasto duplo acontece quando um usuário consegue gastar as mesmas moedas digitais mais de uma vez.

ligação e transmissão de dados entre utilizadores e a rede de confiança e por fim, a aplicação *per si*, a criptomoeda.

Figura 4 - Exemplo das Criptomoedas que se utilizam da *Blockchain*, e a independência entra as camadas.



Fonte: “Tecnologia Blockchain: Fundamentos, Tecnologias de Segurança e Desenvolvimento de Software” (Braga, 2016).

Esta definição funcional das criptomoedas permite concluir tecnologia *blockchain* é independente do funcionamento das criptomoedas e que serve de base de suporte., sendo que para o caso das criptomoedas, a *blockchain* é lastro de confiança, como sugere Ross (2016).

Assim, e por ser uma plataforma tecnológica tecnologia independente em relação aos diversos tipos de criptomoedas, pode ser aplica a outras funcionalidades. Como uma rede distribuída, *peer-to-peer*, e que forneçam respetivamente a confiança e a redução de custo de qualquer aplicação transacional (Ross,2016).

“Após a criação da primeira entrada, todas as alterações ou transações ficam registradas, obedecendo a uma ordem cronológica. Porém, as operações precisam ser validadas antes de serem registradas, criando um elo com a transação anterior” (Katori, 2017), ou seja, um bloco quando escrito com sucesso, não é mais alterado ou apagado.

Desta forma, a *blockchain* é uma base de dados que garante a autenticidade e a segurança de armazenamento de transações sendo que “a tecnologia que sustenta o *blockchain* cria confiança, permitindo que pessoas que não o conheçam (e que, por isso, não têm uma base de confiança subjacente) possam colaborar sem terem de passar por uma autoridade central neutra” (Schwab, 2016).

A confiança na *blockchain*, é sustentada pela imutabilidade dos registos, da alta disponibilidade e dos algoritmos de encriptação. Esta tecnologia é baseada no conceito de *distributed ledger technology (DLT)*¹⁷, que favorece o trabalho em rede (ou seja, em cooperação) como meio de otimizar custos operacionais, mitigar erros ou questões de segurança e melhorar a troca de conhecimento. Para Palatinick (2017) , “o interesse no DLT é impulsionado em grande parte pela necessidade das empresas de aumentar a eficiência operacional e reduzir despesas devido aos desafios contínuos de baixas taxas de juros, restrições de balanço, novos mandatos regulatórios, incertezas macroeconômicas prolongadas e aumento dos custos de conformidade”. (Palatinick, 2017).

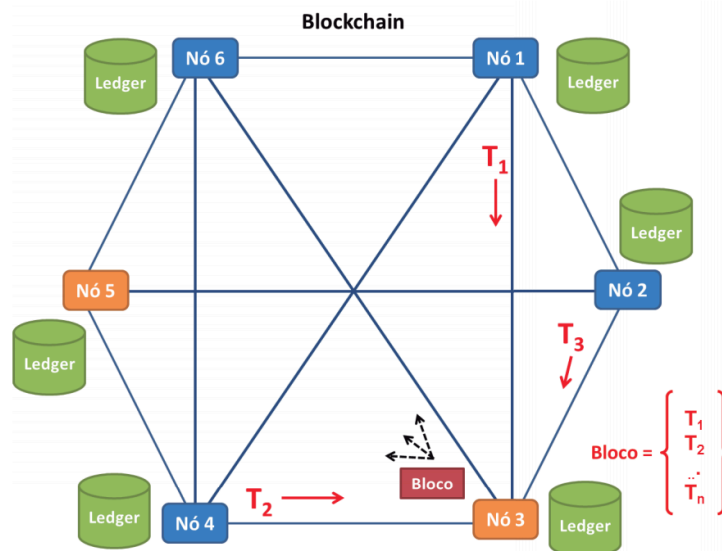
O que é importante ressaltar da afirmação de Palatinick (2017), é o aspeto colaborativo de funcionamento da plataforma, assim como a descentralização sem perder a confiança, o foco e o desempenho de todo o sistema. A descentralização, pode ser conotada como uma melhoria de desempenho para atender o cliente final ou público alvo do serviço, pois permite que o sistema informático esteja mais próximo inibindo situações de congestionamentos ou sobrecargas, eventos característicos de sistemas informáticos centralizados.

Uma rede DLT, é uma rede formada por computadores, que estão ligados entre si e comunicam-se em P2P.

Figura 5 - Exemplo de uma DLT e a comunicação P2P.

¹⁷ Os registos distribuídos usam computadores independentes (chamados de nós) para registrar, compartilhar e sincronizar transações em seus respectivos registos eletrônicos (em vez de manter os dados centralizados como em um *ledger* tradicional). O *Blockchain* organiza os dados em blocos, que são encadeados em um modo somente de anexação. Acedido em 02/05/2018, em <http://www.worldbank.org/en/topic/financialsector/brief/blockchain-dlt>.

A análise preditiva sob o aspecto da regulação.



Fonte: “*Tecnologia Blockchain: Fundamentos, Tecnologias de Segurança e Desenvolvimento de Software*” (Braga, 2016).

A Figura 5 ilustra o modo de funcionamento da plataforma de *blockchain*, formada por diversos computadores, interconectados, e que se comunicam *peer-to-peer*, partilhando os blocos (de forma distribuída). Estes computadores são os chamados “nós”, da rede DLT, que permite o funcionamento da *blockchain*.

Por outro lado, o *blockchain* possibilita criar aplicações que sejam pseudónimas, sendo o *Bitcoin* o exemplo mais conhecido.

A questão do anonimato por parte dos seus utilizadores pode ser uma falácia. Na realidade, “enquanto as chaves públicas de todas as transações, também conhecidas como endereços *Bitcoin*, são registradas na *blockchain*, tais chaves não são vinculadas à identidade de ninguém. Porém, se a identidade de uma pessoa estivesse associada a uma chave pública, poderíamos vasculhar as transações na *blockchain* e facilmente ver todas as transações associadas a essa chave. Dessa forma, ainda que o *Bitcoin* seja bastante semelhante ao dinheiro vivo, em que as partes podem transacionar sem revelar suas identidades a um terceiro ou entre si, é também distinto do dinheiro vivo, pois todas as transações de e para um endereço *Bitcoin* qualquer podem ser rastreadas. Nesse sentido, *Bitcoin* não garante o anonimato, mas permite o uso de pseudônimo” (Emanuel, 2014).

Assim, a *blockchain*, é uma plataforma em rede, com elevado índice de distribuição (de alta disponibilidade) e de funcionamento colaborativo, o que tende a garantir a

A análise preditiva sob o aspecto da regulação.

autenticidade, integridade e imutabilidade das transações, características únicas para o ganho de confiança por parte dos seus utilizadores.

3. Enquadramento da análise preditiva como ferramenta de apoio à decisão

Das atividades mais desafiadoras para as organizações, os processos de tomadas de decisão possuem significativo papel. O rumo ao sucesso, estagnação e fracasso, são resultados das posições tomadas, sendo um fator relevante para sustentabilidade e continuidade de negócio.

O processo de tomada de decisão necessita de informação. Por outro lado, para a criação de valor, esta carece de interpretação, em que a sua análise possibilita detetar padrões, tendências e propicie inovação. Assim, a análise dos dados, possui papéis determinantes.

Para (Angeloni, 2003), “dados, informação e conhecimento são elementos fundamentais para a comunicação e a tomada de decisões nas organizações, mas seus significados não são tão evidentes. Eles formam um sistema hierárquico de difícil delimitação. O que é um dado para um indivíduo pode ser informação e/ou conhecimento para outro”

Salsburg (2009) recorda-nos, que ao longo do século XX, estatística revolucionou a ciência, ao fornecer modelos que permitiram a otimização do processo de pesquisa, e a auxiliar o processo de tomada de decisões nas políticas socioeconômicas.

Nesse sentido, Stigler (1986) afirma que os métodos estatísticos foram desenvolvidos em meio a uma mistura de ciência, tecnologia e lógica para a solução e investigação de problemas em várias áreas do conhecimento.

Para Ignácio (2012) a evolução dos computadores foi o passo decisivo ao popularizar o acesso às técnicas de estatística a todos os investigadores dos mais diversos campos de atuação. Atualmente, os equipamentos e *softwares* permitem a manipulação de grande quantidade de dados, o que veio a dinamizar o emprego dos métodos estatísticos.

Para (Vasconcelos & Barão, 2017), os sistemas informáticos, como grandes protagonistas de sistemas de apoio à decisão, devem estar aptos a: (i) Manipular grandes volumes de dados; (ii) Obter dados a partir de fontes heterogêneas; (iii) Processar os dados obtidos; (iv) Permitir relatórios configuráveis em formatos textuais e/ou gráficos; (v) Efetuar análises e comparações; (vi) Suportar simulações em função de cenários hipotéticos.

Nos anos de 2008 e 2009 foram produzidos cerca de 6,3 quintiliões de bytes todos os dias e surpreendentemente 90% dos dados do mundo foram criados nos últimos anos,

decorrente a adesão das grandes empresas à internet e criação de redes sociais e dispositivos móveis (Hilbert & Lopez, 2011).

Em síntese, a tecnologia de *big data* proporciona um melhor aproveitamento e utilização das técnicas de AP.

Para Lindstrom (2017, p. 28 e p.29) ilustra o acompanhamento de tendências comportamentais de consumidores de um determinado restaurante americano. Assim, para este autor “a análise de imagens de em vídeo a partir do ano 2000, um proprietário de um restaurante de Nova Iorque que publicou um estudo anonimamente na Craigslist¹⁸, estimava que em 2004 os clientes passavam uma média de 65 minutos à mesa, número que aumentou para um hora e 55 minutos em 2014. Em 2004, as pessoas iam a um restaurante e, de uma amostra de 45 pessoas, três pediram para mudar de mesa. O grupo de amostra, passou uma média de oito minutos a decidir o que ia pedir. As entradas que pediram foram servidas ao fim de seis minutos. Dois dos 45 clientes mandaram para trás a comida, queixando-se que estava fria. A média dos clientes saía do restaurante cinco minutos depois de pagar a conta.

Uma década mais tarde, as coisas mudaram. Atualmente, 18 em 45 clientes que entram num restaurante pedem para mudar de mesa. A partir daí, as suas vidas digitais assumem o comando, sendo que os clientes pegam nos telemóveis e tentam ligar-se ao *wi-fi* mais próximo e procuram informação ou veem se alguém *gostou* da sua publicação no Facebook, frequentemente esquecendo que a sua ementa está à espera em cima da mesa, e é por isso que quando o empregado de mesa lhes pergunta se estão prontos a pedir, a maioria responde que precisa de mais tempo. Vinte e um minutos mais tarde, estão prontos para pedir. Vinte e seis deles passam até três minutos a tirar fotografias da comida. Catorze tiram fotografias uns dos outros a comer, e se as fotografias ficam tremidas ou mal, voltam a tirá-las. Aproximadamente metade dos clientes pergunta se o empregado se importa de tirar uma fotografia do grupo e já agora se não se importa de tirar mais algumas? A segunda metade manda a comida para trás, dizendo que está fria (e está, uma vez que passaram dez minutos a brincar com os telemóveis e não a comer). Depois de pagarem a conta, saem do restaurante vinte minutos mais tarde, por oposição aos cinco minutos de demoravam em 2004. Quando estão a sair, oito dos clientes estão tão distraídos que chocam uns com os outros, com um empregado ou com uma mesa ou

¹⁸ O site da *Craigslist* possui fóruns sobre diversos assuntos.

cadeira (Lindstrom, 2017).”

Esta análise só foi possível através das imagens capturadas, da correlação de todos os acontecimentos e métrica. Assim, com uma análise afinada, a direção do restaurante possui os elementos essenciais para moldarem o negócio aos clientes: a mudança de comportamento dos clientes.

Este exemplo sugere que a competitividade e continuidade das atividades de uma organização passam por um processo contínuo de avaliação e adequação e até mesmo antecipação,

“A procura de criação de valor para o acionista obriga à concentração de competências, à seleção de negócios essenciais, à adoção de medidas de otimização de recursos, de *downsizing* e à eliminação de níveis hierárquicos – visando tornar-se mais ágeis e produtivas, focalizando-se no que melhor sabem fazer” (Carvalho, 2014).

Portanto a prática do popular jargão “fazer mais com menos”, depende de um processo contínuo de adequação do negócio, em que permita avaliar diversos contextos e favorecer tomadas de decisões com precisão. Num mundo globalizado, velocidade é essencial, pelo que a tomada de decisão não pode ser um produto de “lenta decisão”; nem negligente, deve analisar de forma segura as possibilidades e riscos.

De forma complementar a afirmação anterior, os recursos tecnológicos no âmbito da AP e *big data*, facilitam “a definição do perfil de utilizador através de técnicas de análise e inferência de grandes volumes de dados está a abrir caminho para novos serviços, muito mais personalizados, que podem beneficiar utilizadores e consumidores, mas que também levantam preocupações importantes relativamente à privacidade e autonomia individual dos utilizadores” (Schwab, 2016).

A AP é, também, um pilar da evolução mais acelerada da inteligência artificial, sendo que o tratamento de dados em tempo real permite tomadas de decisões, também, em *near real-time*.

Nesse sentido, Nadella (2018, p. 148) defende que “a inteligência artificial é o motor de todas as experiências, aumentando a capacidade humana com resultados e poderes preditivos que seria impossível alcançarmos sozinho”

3.1 A análise preditiva como ferramenta de apoio à tomada de decisões

As organizações tomavam suas decisões com base no conhecimento e experiências de especialistas, que de certa forma, causava influências nas operações do dia a dia. As técnicas estatísticas surgiram com a intenção de descobrir padrões de dados invisíveis a percepção. Em dias atuais, em que capturamos dados em um volume cada vez maior, como podemos verificar em publicação da DOMO (ver item 2.2), estas técnicas estão se tornando indispensáveis para extrair valor a partir destes dados. A analítica é capaz de produzir estatísticas e previsões confiáveis (Guazzelli, 2012).

A AP, conjuntamente com modelos de previsões e técnicas de mineração de dados, aplicados em uma plataforma de *Big Data*, dependem de sofisticados métodos de estatísticas, de diversos tipos, como os modelos de regressão ou series temporais avançadas. Essas técnicas permitem às organizações determinarem tendências e relações que podem não são de fácil percepção, pois o manancial de dados é extremamente grande, e correlacionar é uma árdua tarefa para o ser humano.

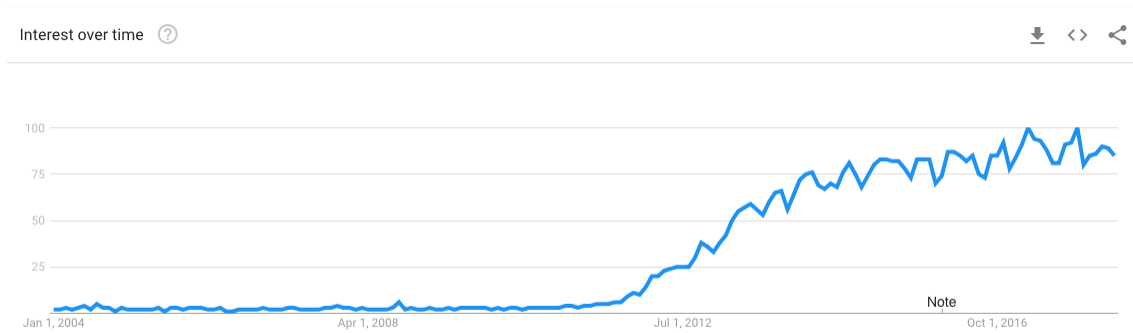
Desta forma, a AP permite prever eventos ou comportamentos futuros, facto determinante para a inovação e sustentáculo da competitividade.

Para (Ross, 2016), “*Big data* é uma expressão genérica que é utilizada para descrever a forma como estes grandes volumes de dados podem ser atualmente utilizados para compreender, analisar e prever tendências em tempo real. Esta expressão pode ser usada indiscriminadamente para significar *análise de big data*, *análise de dados* ou *análise de dados aprofundada*. Um mal-entendido frequentemente consiste em pensar que os avanços possibilitados pelos *big data* dependem simplesmente da quantidade de dados reunida. Na verdade, o crescimento do volume de dados sem a capacidade de os processar não é útil por si só”

Big Data e AP, são, pois, dois pilares de transformação e que estão a inovar o entendimento comportamental das pessoas e grupos de pessoas, e que habilita as organizações a serem mais competitivas.

O interesse por *big data*, despontou entre 2011 e 2012, tempos que uma das maiores redes sociais, ultrapassou a barreira de um mil milhões de utilizadores. Época em que o despertar de interesse pelo que a análise comportamental de utilizadores passou a estar de forma avassaladora presente.

Figura 6 - Nível de interesse por *Big Data* em pesquisas pelo item. Medido mediante pesquisas na página da Google.



Fonte: <https://trends.google.pt/trends/explore?date=all&q=%2Fm%2F0bs2j8q> Acedido em 20/05/2018.

Não apenas as redes sociais, sendo estas talvez, o princípio desta abundância de dados e informação, porém “existem áreas de I&D relacionadas que potenciam o crescimento dos dados como, tais como a IoT¹⁹, a evolução das tecnologias de *Data Mining*, no sentido do *Big Data*, e as tendências da tecnologia industrial, com uma nova revolução designada por Indústria 4.0” (Vasconcelos & Barão, 2017, p. 29 e p.30).

A IoT, permite atribuir sensores e obter dados dos mais variados, como *smartbands*, que recolhem informações pessoais (como batimentos, número de passos por dia, horas de sono).

São imensos os exemplos. Existe balanças que se ligam aos *smartphones* e de uma forma prática consegue sincronizar dados e manter um histórico, aplicações que registam nosso hábito alimentar, e muitos outros exemplos. E no setor industrial, os sensores que estão a captar todos os dados, dos setores produtivos e consumo final, permite a indústria ter o total conhecimento do seu funcionamento e permitir a otimização de processos e custos.

Para Schwab (2016, p. 10 e p.11), “a quarta revolução industrial, contudo, não se refere apenas a máquinas e sistemas inteligentes e conectados. O seu alcance é muito mais vasto. Ondas de novas descobertas ocorrem simultaneamente em áreas que vão desde o sequenciamento genético à nanotecnologia, passando pelas energias renováveis ou pela computação quântica. É a fusão destas tecnologias e a interação entre os domínios físico, digital e biológico que torna a quarta revolução industrial radicalmente diferente das

¹⁹ *Internet of things*

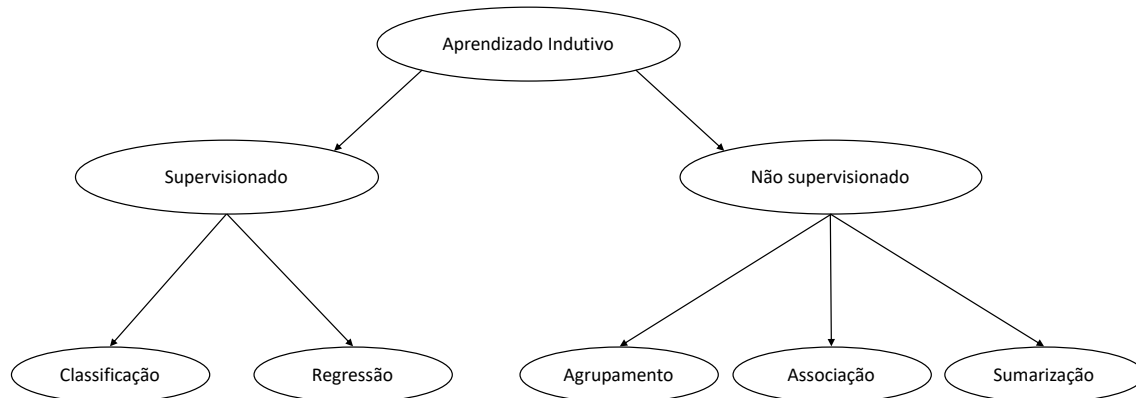
revoluções anteriores.”

A AP está na base das diversas tecnologias disruptivas, estão a ajudar a compreender e a moldar o mundo – “*Prediction is booming. It reinvents industries and runs the world*” (Siegel, 2016). O termo “aprendizado de máquinas” tem se popularizado.

“A área de aprendizado de máquina é considerada um ramo da Inteligência Artificial, cujo objetivo é construir sistemas que sejam capazes de aprender de forma automatizada a partir de dados. Algoritmos que aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir desses dados que representam” (Mota & Siqueira, 2016), sendo, pois, uma das áreas da inteligência artificial (IA).

Estes tipos de algoritmos, “são utilizados em diversas tarefas, que podem ser organizadas de acordo com diferentes critérios. Essas tarefas de aprendizado podem ser divididas em descritivas e preditivas. Nas tarefas descritivas, a meta é explorar ou descrever um conjunto de dados. Já nas tarefas preditivas a meta é encontrar uma função a partir dos dados de treinamento que possa ser usada para prever valores desconhecidos ou futuros de outras variáveis” (Mota & Siqueira, 2016).

Figura 7 – Hierarquia de aprendizado.



Fonte: (Faceli, Lorena, Gama, & Carvalho, 2011)

Através do denominado aprendizado indutivo são realizadas generalizações a partir dos dados, sendo que os tipos de aprendizado supervisionado e não supervisionado correspondem, respetivamente, aos modelos preditivo e descritivo.

As tarefas supervisionadas diferenciam-se pelo tipo de dados: (i) no caso de classificação, discreto; (ii) no caso de regressão, contínuo. Já as tarefas não supervisionadas são

divididas em agrupamento, associação e sumarização. No agrupamento, os dados são agrupados de acordo com a similaridade; na associação, o objetivo é encontrar padrões frequentes de associações entre os atributos de um conjunto de dados; já a sumarização consiste em encontrar uma descrição simples para um conjunto de dados.

“*Machine Learning* é um tópico de investigação na área de IA para o desenvolvimento de mecanismos de aprendizagem com base em dados provenientes de exemplos, experiências e fenômenos em geral” (Vasconcelos & Barão, 2017, p. 246).

“Um dos 4 quadros mais antigos e mais usados para descrever o percurso do consumidor é o AINDA: *atenção, interesse, desejo e ação*” (Kotler, 2017, p. 86). A AP pode reforçar os trabalhos das organizações para descobrir o que chama a atenção, através de *tracking* comportamental nas plataformas digitais, revelar o que realmente interessa e deseja o consumidor. São passos de extrema importância em um mundo que a compra está a um clique de distância, as organizações necessitam agir de forma rápida, antecipada se possível e no momento certo.

3.2 Casos de aplicação de análise preditiva

Algumas práticas de AP, assente em *Big Data*, demonstram diversas mais-valias e aplicabilidades nos mais diversos setores.

“Em eleições muito competitivas, a campanha de Obama utilizou *big data* para obter informações sobre como angariar dinheiro, os locais onde fazer campanha e como publicitar, uma estratégia com a qual nenhum dos seus adversários conseguiu rivalizar” (Ross, 2016). O autor afirma, que na campanha de 2012, Obama conseguiu angariar 690 milhões de dólares a partir de doações *online*, num total de 1.123.000 de dólares. Uma performance aproximada de duas vezes superior a campanha republicana de Mitt Romney.

Para outro elemento da campanha, e também responsável pelo tratamento da informação, Michael Slaby, afirma que o termo “*big data* significa na verdade é a capacidade de processar imensa informação num período que se aproxima do tempo real, para podermos fazer efetivamente alguma coisa com ela” (Ross, 2016, p. 171). Sendo o termo *big data*, para Slaby, uma síntese da capacidade de armazenamento e AP de dados, sendo áreas tecnológicas complementares.

A importância de observar os dados e extrair a informação de valor correta pode redefinir o futuro de uma organização.

Analisar tendências e padrões é importante, mas inovação ou reinvenção pode estar “debaixo das menores pedras”. Como no caso da LEGO, em que no seu processo de tomada de decisão, “baseava-se inteiramente em grandes quantidades de megadados. No entanto, em última análise, foi uma pequena informação acidental – um par de tênis pertencentes a um *skater* e apaixonado por legos – que ajudou a impulsionar a recuperação da empresa. A empresa não só redesenhou os blocos para o seu tamanho original, como começou a adicionar mais e menores blocos às caixas. Os blocos tornaram-se mais detalhados, os manuais de instruções mais pormenorizados, os desafios de construção mais trabalhosos” (Lindstrom, 2016, p. 9).

A AP deve ser capaz de analisar o grande volume de informação sem deixar escapar a “pequena” informação, que de modo geral poderia ser descartada, pela sua baixa representatividade.

Para a Tesla, utiliza das atualizações de *software* e as ligações *wireless* para melhorarem seus produtos. “A análise disponibilizada pelos sensores colocados nos produtos permite a sua monitorização e a manutenção proactiva e, ao fazê-lo, maximiza a sua utilização. Já não se trata de encontrar falhas específicas, mas sim de usar referências de desempenho (com base nos dados fornecidos por sensores e monitorizados através de algoritmos) que podem alertar quando uma parte do equipamento se afasta do seu funcionamento normal” (Schwab, 2016, p. 53 e p.54). A Tesla também se beneficia da AP através do aprendizado de máquina para o desenvolvimento da condução autónomas em que o carro necessita tomar decisões e antever situações, sendo possível ao tratamento de dados e em tempo real.

O tratamento da informação, com apoio das tecnologias de *big data* e AP são partes de uma crescente onda de novas tecnologias que estão a proporcionar a disrupção em diversos mercados.

Neste sentido, Carvalho (2014, p.179) sugere que “há um reconhecimento crescente de que as tecnologias periféricas de uma empresa são geralmente, as atividades-chave de outras”

Na realidade. E apesar da tecnologia não ser produto ou serviço fim de uma organização, claramente possui sua utilidade, se bem compreendida e praticada.

Em outro exemplo, “Wall Street tirou tanto partido dos *big data* como qualquer outro sector de atividade. Dos cerca de sete mil milhões de ações transacionadas nos mercados de capitais dos EUA todos os dias, dois terços são negociados por algoritmos informáticos previamente programados que trituram dados sobre preços das ações, o melhor momento para negociar e a quantidade para maximizar os ganhos e minimizar o risco. Chama-se a isto negociação algorítmica, sendo atualmente a norma no mundo financeiro” (Ross, 2016, p. 182 e p. 183).

A área da banca, talvez seja o setor que esteja a passar pelo seu momento disruptivo. “A aplicação dos *big data* para melhorar as operações e o desenvolvimento de produtos na banca a retalho é conhecida como *fintech*”, complementa Ross (2016, p. 183).

Um caso interessante da aplicação de AP em um grande volume de dados, foi proporcionado pela Google, no documentário *The Human Face of Big Data*, em que a empresa conseguiu antecipar epidemias de gripe até duas semanas antes do que os sistemas informáticos do sistema de saúde americano. “Até recentemente, a única maneira de detetar uma epidemia de gripe era acumulando informações enviadas pelos médicos sobre as visitas dos pacientes” (Smolan & Erwit, 2013).

4. Modelo de Investigação

O principal objetivo de um trabalho de investigação, a desenvolver neste capítulo, é o foco na decisão sobre o tipo de enquadramento do tema que se deseja desenvolver (contextualização e importância da análise preditiva e da garantia dos direitos e deveres do Regulamento Geral de Proteção de Dados da Comunidade Europeia) com o mercado competitivo, que possibilita uma investigação consequente.

Sendo atual o tema da tese, a análise dos diversos estudos, carecem de interpretação e conexão. Para tal, é importante que “interpretação se faça a partir da ligação dos dados com conhecimentos significativos, originados de pesquisas empíricas ou de teorias comprovadas” (Gil, 1991).

O método indutivo se inicia em algo particular e segue para algo mais amplo. “Indução é um processo mental por intermédio do qual, partindo de dados particulares, suficientemente constatados, infere-se uma verdade geral ou universal, não contida nas partes examinadas. Portanto, o objetivo dos argumentos indutivos é levar as conclusões cujo conteúdo é muito mais amplo do que o das premissas nas quais se basearam” (Lakatos & Marconi, 1991).

4.1 Recolha de Informação

Geralmente os estudos de caso baseiam-se em seis fontes principais: documentação, registos de arquivo, entrevista, observações diretas, observações participantes e factos físicos (Yin, 1994).

O tema *blockchain*, em tempos atuais, aparenta com um tópico de discussão recente e com algum dinamismo, o que sugere novos caminhos que podem solucionar questões da sociedade atual, como a privacidade *versus* sustentabilidade da competitividade que as empresas retiram aparentemente do tratamento e análise de dados.

4.2 Perguntas de partida

Quivy e Campenhoudt (2013) propõem que o problema de investigação tende a dar origem a uma pergunta ou sugestão, onde o investigador tenta exprimir com precisão o

que procura saber, elucidar e compreender melhor, pelo que “uma boa pergunta de partida deve exigir clareza, exequibilidade e pertinência”.

A questão de partida centralizou-se na análise do processo de tratamento de dados relativos a pessoas, de certa forma popularizada pelas redes sociais, e do conflito com o recente RGPD.

Assim, o presente trabalho de investigação tem como perguntas de partidas que servem de norte:

Q1 A análise preditiva é relevante para as organizações?

A evolução dos sistemas informáticos associados com as técnicas de análise de dados tem otimizado a gestão das organizações, nas tarefas mais simples, ou até mesmo na automação de processos (exemplo: automação industrial). Desta forma, a utilização de dispositivos móveis e o crescimento de adoção às redes sociais massificou os rastros de dados que as pessoas deixam na internet.

“A presença de um indivíduo mede-se pelas suas interações digitais e pelos vestígios que deixa numa multiplicidade de plataformas e meios *online*” (Schwab, 2016). Este fenómeno parece alimenta a necessidade que as organizações precisam, que é conhecer o mercado e oferecer produtos/serviços atrativos. Pelo que parece que empresas que conseguiram tirar partido destas informações, tenham uma nova arma para melhorar a competitividade.

Q2 Qual a importância da RGPD nas organizações?

A proteção de dados é um tema que preocupa a Europa desde o fim da II Guerra Mundial, desde a criação do Conselho da Europa (CdE) e que “reuniu 10 países europeus com o objetivo de promover o Estado de direito, a democracia e os direitos humanos” (Saldanha, 2018).

A digitalização melhorou substancialmente o acesso a informação e também a comunicação. Por outro lado, também os ataques informáticos estão sempre a abalar a confiança destes sistemas, o que preocupa a sociedade com os direitos individuais.

O Regulamento Geral de Proteção de Dados (RGPD) é um avanço importante para o equilíbrio da sociedade.

Q3 A tecnologia *blockchain* pode assegurar a realização de tratamento de dados face ao RGPD?

A avaliação de soluções que permitam aumentar o nível de confiança da sociedade em relação à proteção de dados, e em tempos de digitalização, é um tema proeminente. A confiança é um dos ingredientes fundamentais para a atribuição de valores. Em um mundo globalizado, a competitividade não se restringe ao local é importante saber o que o cliente quer e quando quer. As redes sociais, como Facebook, possuem valiosas informações, como: o que uma pessoa gosta, por onde anda, onde passa férias e etc. Todos estes vestígios no mundo *online* preocupam a sociedade, mas agradam as empresas.

Tecnologias como o *blockchain*, que na prática é uma base de dados dinâmica, e que tende a fornecer confiança, nunca sofreu um ataque informático com sucesso, como afirma Alec Ross (2016). Estes dois ingredientes podem indicar um caminho para a temática da privacidade e proteção dos dados, mas sem retirar a capacidade das empresas em analisar dados. É através da AP que muitas empresas melhoram seus produtos, ou descobrem novas oportunidades.

4.3 Instrumentos

Este trabalho assenta na captura, observação, análise e tratamento de dados recolhidos. “Na observação, são aplicados atentamente os sentidos a um objeto, a fim de que se possa, a partir dele, adquirir um conhecimento claro e preciso. A observação deve ser exata, completa, imparcial, sucessiva e metódica” (Kauark, Manhães, & Medeiros, 2010).

Sendo uma observação sistematizada, “também denominada observação planeada, é estruturada e realizada em condições controladas, de acordo com objetivos e propósitos previamente definidos” (Kauark et al., 2010).

“De acordo com Aristóteles, no primeiro momento, devia-se iniciar pelo que vinha em primeiro lugar no conhecimento, que seriam os fatos percebidos pelos sentidos e, depois, agrupar as observações, pelo processo de indução, em uma generalização que proporcionasse a forma universal, isto é, a substância, a identidade inteligível e real que permanecia independente das mudanças. O objetivo desse processo indutivo de abstração, e da ação do entendimento agente iluminador, era o de definir as formas e efetuar a

passagem progressiva dos dados materiais e mutáveis para os imateriais e imutáveis” (Paulo & Cesar, 2009).

A recolha de informações, em contexto com os acontecimentos da análise e respetivo tratamento de dados, da necessidade de garantia e privacidade dos indivíduos e a necessidade de manter a competitividade das empresas (daquelas que se utilizam da AP como principal ferramenta do processo de decisão), são os temas de análise deste trabalho académico.

5. A Análise Preditiva após a implantação do Regulamento Geral de Proteção de Dados

As tecnologias de *big data* e AP são melhor entendidas, como uma versão mais poderosa da descoberta de conhecimento em bancos de base de dados ou mineração de dados. A mineração de dados permite que as empresas descubram inferências de fatos e padrões previamente desconhecidos em uma base de dados e que se traduz em resultados para aquelas organizações que dispõem das tecnologias citadas.

O regulamento introduz alterações destinadas a reforçar o controlo individual dos dados pessoais, reforçando as disposições relativas à transparência e ao consentimento, revendo a disposição que prevê e anunciando novos direitos individuais. Além disso, amplia as responsabilidades dos detentores dos dados, por meio de uma série de novas disposições de prestação de contas, como indica Rubinstein (2013) e complementa:

“Article 22 summarizes the responsibilities of controllers including general obligations such as documentation (Article 28), data security (Article 30), impact assessments (Article 33), prior authorization or consultation (Article 34), and designation of a data protection officer (Article 35). In addition, Article 23(2) creates a more specific obligation for controllers to implement mechanisms to ensure, by default, that data minimization requirements are satisfied” (Rubinstein, 2013).

“Não é a primeira vez que direitos ligados à personalidade, como o direito de privacidade, são colocados em xeque diante da evolução tecnológica. Isto se demonstra pelo famoso episódio que substancialmente inaugurou o debate sobre o direito à privacidade, em nível mundial: tratava-se do surgimento das primeiras máquinas fotográficas instantâneas e seu uso pela chamada imprensa marrom, em 1890, o que motivou o ensaio de Warren e Brandeis sobre *The Right to Privacy*, consagrando o direito de estar só” (Colombo & Neto, 2017).

Em recentes ataques informáticos, proporcionados por ameaças conhecidas como *WannaCry*²⁰ ou *Meltdown*²¹, que provocou a fuga de informações armazenadas pelas

²⁰ “*Wannacry*, consistiu na disseminação de um software malicioso, que bloqueia os computadores com sistema operativo Windows e exige um pagamento para os libertar.” (Jornal de Negócios, 2017).

²¹ “O *Meltdown* e o *Spectre* são falhas informáticas que permitem a terceiros aceder a informação confidencial e sensível, como passwords. O *Meltdown* afeta os computadores com processadores Intel, AMD e ARM.” (DECO, 2018).

organizações. O caso que ocorreu com a *Uber*²² é um exemplo impactante de quebra de segurança. Conforme foi amplamente publicado pela mídia, “a plataforma digital de transporte Uber pagou cem mil dólares (85 mil euros) aos hackers que piratearam dados de 57 milhões de utilizadores e condutores da empresa e, na sequência do ataque informático, despediu o responsável pela área da segurança.” (Notícias, 2017).

Nesse contexto, “os desafios culturais são enormes e, é claro, as preocupações com a privacidade só se tornarão mais significativas. Mas as tendências subjacentes, tanto na tecnologia quanto no retorno do negócio, são inconfundíveis” (Brynjolfsson & McAfee, 2012).

Paradoxalmente, por um lado é necessário adaptar a tecnologia, para que os direitos a dignidade humana e a preservação da individualidade não seja banalizada, porém é importante assegurar as conquistas e as capacidades analíticas que vem vindo a ocorrer.

Assim, “os níveis de confiança na sociedade oscilam em função do comportamento dos governantes, das instituições, das decisões tomadas por outros membros da sociedade, entre outros aspectos” (Correia, 2017, p. 30).

5.1 *Blockchain* e sua multifuncionalidade

A criptomoeda mais conhecida, a *bitcoin*, utiliza tecnologia *blockchain* como suporte ao seu funcionamento e possui como regra básica a segurança e a privacidade. Deste modo “os endereços de rede *bitcoin* são representados por chaves públicas, ou mais concretamente, por derivações de *hash*²³. Estes endereços são anónimos porque não contêm elementos identificativos dos seus proprietários como, por exemplo, nomes ou números de identificação nacionais” (Martins, 2018, p.89).

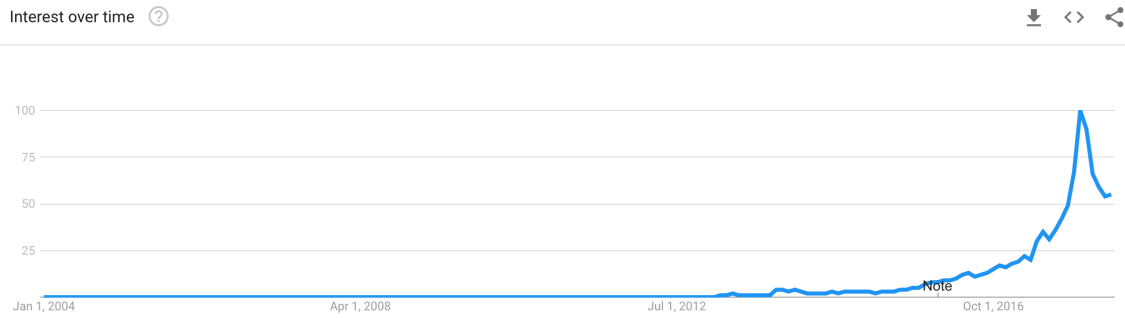
O interesse pelo tema *blockchain*, como se pode aferir no gráfico seguinte, com ajuda da ferramenta Google Trends, subiu consideravelmente nos últimos tempos, com as notícias da valorização do *bitcoin*, mas que evidencia a tecnologia.

Figura 8 - Google Trends - Nível de interesse por *Blockchain* em pesquisas pelo item. Medido mediante pesquisas na página da Google.

²² Famosa plataforma digital que ligam pessoas que se querem deslocar na cidade, e pessoas disponíveis para as levar onde querem ir.

²³ Um *hash* (ou escrutínio) é uma sequência de bits geradas por um algoritmo de dispersão, em geral representada em base hexadecimal.

A análise preditiva sob o aspecto da regulação.



Fonte: <https://trends.google.pt/trends/explore?date=all&q=%2Fm%2F0138n0j1>, acessado em 31/05/2018.

Desta forma é previsível que “o *blockchain* poderá proporcionar uma solução de custo muito mais reduzido para as transações que exigem um intermediário terceiro como avalista, como documentos legais, comissões de corretagem e compra de bilhetes” (Ross, 2016, p.128). Uma clara perceção que é possível aplicar a utilizações de outras funcionalidades, não apenas na criptomoeda, da segurança, confiança e garantida privacidade que a tecnologia *blockchain* proporciona.

Segundo Ross (2016, p.129), Joi Ito, diretor do MIT Media Lab, considera “o *blockchain* será para o sector bancário, jurídico e de contabilidade aquilo que a Internet foi para os meios de comunicação social, comércio e publicidade. Irá reduzir os custos, eliminar intermediários em muitas camadas da atividade e reduzir o atrito. Como sabemos, o atrito de uma pessoa é o rendimento da outra”, afirma.

A área de saúde, possui dados caracterizados como especiais pelo artigo 9º do regulamento, “pela sua natureza, são especialmente sensíveis do ponto de vista dos direitos e liberdades fundamentais, pelo que o tratamento desses dados pode implicar riscos significativos para esses mesmos direitos e liberdades” (Saldanha, 2018, p.35).

Garantir a privacidade, como direito fundamental, mas permitir o tratamento da informação para extração de métricas e informações de valores é um proeminente desafio.

“Unlike ID cards or conventional medical records, blockchains promise selective data sharing through adequate applications, ensuring privacy and reducing the risk of identity theft. Blockchains could thus facilitate new forms of identity management by enabling

individuals - to control access to their identity information and to create, manage and use a self- sovereign identity” (Fink, 2017).

A Autoridade Europeia para a Proteção de Dados reconhece que “embora as tecnologias avançadas aumentem o risco à privacidade e proteção de dados, elas também podem integrar soluções tecnológicas para uma melhor transparência e controle para as pessoas cujos dados são processados” (Fink, 2017).

Neste enquadramento, o *blockchain*, é sugerido como exemplo de plataforma tecnológica que permite o armazenamento de dados, tratamento e processamento, que gera confiança e transparência. Atributos desejáveis pelas organizações, no cumprimento da regulação, e por parte dos clientes, que avaliam a forma que seus dados são utilizados, terem o controlo e garantir o direito a privacidade é essencial.

Uma recente divulgação de informação sobre determinado tema, pode ser compilada através de uma série de aplicações em escala mundial da tecnologia *blockchain*, e em diversos setores de atuação. O infográfico abaixo, permite-nos ter um panorama das diversas utilizações. “*A number of governments have expressed an interest in blockchain technology to store public records on a decentralized data management framework. Essentia²⁴ is developing an e-government pilot with Finland’s Central Union of Agricultural Producers and Forest Owners. Blockchain will enable urban and rural citizens throughout Finland to access records. Other use cases include government applications such as education, public records and voting” (Zago, 2018).*

Figura 9 – Infográfico: Diversas utilizações para o *Blockchain*

Controlo de Fronteiras A Holanda utiliza a plataforma para armazenar dados de passageiros.	Notários O estado Americano da Georgia, utiliza a tecnologia para registar as propriedades. Desenvolvida pela <i>National Agency of Public Registry</i> .
Venda de Diamantes A empresa Dee Beers, utiliza a tecnologia para rastrear a importação e venda de diamantes.	Saúde Existem implementações pioneiras para o armazenamento dados médicos e de saúde em <i>blockchain</i> , como a implementada pela empresa MedRec.
Cadeia de Suprimentos A america Walmart, em parceria com a IBM, criaram um projeto para o monitoramento de alimentos na China, para garantir a qualidade.	

Adaptado da fonte: <https://medium.com/@matteozago/50-examples-of-how-blockchains-are-taking-over-the-world-4276bf488a4b> , Acedido em 02/06/2018.

²⁴ Essentia é uma estrutura modular, que permite aos usuários operar com suas identidades e dados descentralizados, com recursos *on-chain* que podem ser acedidos de todos os dispositivos do mundo.

Nos EUA, alguns projetos estão a experimentar a plataforma tecnológica de *blockchain* para “transações imobiliárias, de compra, venda, hipoteca, aluguel, registros de novas terras e de demolições ficarão armazenados num sistema *blockchain*. Além disso, espera-se que serviços notariais e de contratos inteligentes também serão inseridos no sistema. Da mesma forma, a Ucrânia assinou uma parceria para implementar o *blockchain* nas mais diversas áreas estatais. Outros países com iniciativas semelhantes são Estônia e Suécia” (Katori, 2017).

A *blockchain* é um meio tecnológico que proporciona segurança, um fator essencial para a confiança necessária, sendo uma aposta segura para adoção pelas organizações para diversas funções. Assim como Ross (2016), outro autor afirma que “a *bitcoin* corre em todo o mundo, 60 segundos por minuto, 60 minutos por hora, 24 horas por dia, 365 dias por ano, há 9 anos consecutivos sem nunca ter sofrido um ataque bem-sucedido. E não foi por falta de ser atacada” (Pacheco, 2018, p. 85).

5.2 *Blockchain*, análise preditiva e o Regulamento Geral de Proteção de Dados

Em mundo digital, com consumidores bem informados “um único momento de deleite inesperado de uma marca é o que basta para transformar um consumidor num defensor fiel de uma marca. Para isso, as empresas têm de mapear o percurso do consumidor até à compra, compreender os pontos de contacto do consumidor ao longo do percurso e intervir nos pontos de contacto selecionados e importantes. Devem concentrar os seus esforços – intensificação das comunicações, reforço da presença do canal e melhoramento da interface com o consumidor – para melhorar esses pontos de contacto fundamentais e introduzir uma forte diferenciação” (Kotler, 2017, p. 82). É importante ressaltar que “a privacidade está relacionada à vida pessoal e a proteção de dados diz respeito às informações pessoais” (Fabiano, 2018).

Para que as organizações possam realizar este trabalho, é preciso obter dados das pessoas e mapear seu perfil. E a partir do RGPD estes dados tem que ser utilizados com transparência e devem estar regidos pelo regulamento pelo que as pessoas necessitam confiar na organização para permitir a partilha de informação.

“Confiamos no nosso médico, obrigado a sigilo profissional normalmente respeitado. Mas acontece o mesmo com as informações confidenciais que revelamos nas páginas da Internet que visitamos e nas redes sociais? A questão da confidencialidade é tão vívida para nossas interações digitais como para os dados médicos, mas as nossas garantias na matéria são muito inferiores” (Tirole, 2018, p. 395).

Para entender as implicações da adoção de *blockchains*, a partir do ponto de vista privacidade, devemos aprofundar um pouco mais nos seus detalhes técnicos.

Numa rede de dados *blockchain* os dados são agrupados em blocos que, ao atingir um determinado tamanho, são encadeados a cadeia existente por meio de um processo de *hashing*. “De maneira simplificada a *Blockchain* é uma estrutura de dados que armazenatransações de forma ordenada e ligada ao bloco anterior, servindo como um sistema de registros distribuído” (Chicarino, Ferreira Jesus, Albuquerque, & Rocha, 2017).

“Blockchain-enabled applications involving security and privacy sensitive data, e.g., financial and healthcare, requires confidentiality, security and privacy assurances at different levels to be supported by the system, which are mandated by regulatory compliance requirements such as HIPAA²⁵ and GDPR” (Tam Vo, Kundu, & Mohania, 2018).

Com a *blockchain* é puramente um repositório de transações, um mecanismo de execução será necessário para análises executadas diretamente aos dados. Uma das soluções possíveis para esse problema, é tornar os dados *blockchain* prontamente acessíveis por sistemas de processamento paralelo de dados, com métodos de AP.

Há também casos de uso, como aplicativos IoT, como aponta (Tam Vo et al., 2018). Este tipo de *approach* permite uma análise em tempo real. A *blockchain* ou “cadeia de blocos também pode ser usada para comunicações em cadeia de fornecimento, contratos inteligentes, gerenciamento de identidade digital e em uma série de outras aplicações” (Pilkington, 2015).

O ponto de partida é que todos os participantes numa rede *blockchain*, na qual os dados pessoais são armazenados, são responsáveis por esse armazenamento e processamento

²⁵ Aprovada em 1996 a lei de portabilidade e responsabilidade de seguro saúde - *Health insurance portability and accountability act*.

sob o RGPD. Se as partes que executam nós da rede DLT, ou seja, elementos que armazenam dados e os processam, forem realmente os participantes (responsáveis), então serão enquadrados no art. 28 do RGPD, obrigados a manter acordos com as partes que executam a função de nó da rede DLT. Isso não é muito prático (pois provavelmente exigiria uma quantidade significativa de acordos para tratar os dados), mas, em teoria, certamente não é inviável.

A *blockchain* por causa de sua inalterabilidade, entra em conflito com o RGPD, pois assim que os dados pessoais não forem mais necessários, eles devem ser apagados e o titular dos dados tem o direito de solicitar o apagamento. Da mesma forma, os titulares de dados têm o direito de alterar dados pessoais incorretos ou desatualizados. Em relação ao primeiro - o apagamento de dados pessoais - isto poderia ser conseguido, ao invés de apagar os dados, criptografando os dados pessoais e excluindo a chave usada posteriormente. Dessa forma, os dados (originais) não seriam removidos, mas um bloco (extra) seria adicionado detalhando a criptografia (e a subsequente exclusão da chave).

Da mesma forma, se as partes que processam os dados (nós), forem consideradas agentes de tratamento de dados à luz do RGPD, em vez de concluir acordos com vários *processadores* baseados no art. 28 RGPD com as partes que executam nós, esta questão pode ser facilmente resolvida através de acordos conjuntos entre todos os participantes e partes que executam os nós. Em relação ao terceiro problema, o apagamento de pessoal dados, isso pode ser facilmente resolvido, pois a remoção de dados pessoais requer a maioria de todos os nós (com base no resultado da maioria das pesquisas).

Isso significa que num *blockchain* opaco, em que uma parte tem o poder de mais de 50% + 1 de todos os nós, o apagamento de dados (pessoais) da *blockchain* é muito viável. Nesse cenário, a maioria dos nós apagaria os dados e todos os outros nós também os apagaríamos posteriormente. “Como tal, a obrigação de apagar dados pessoais do RGPD não impede o armazenamento de dados pessoais em um *blockchain*. Em relação ao quarto problema, a alteração de dados pessoais, isso poderia ser facilmente resolvido da mesma forma que o apagamento de dados, alterando os dados armazenados para a maioria de todos os nós” (Geelkerken & Konings, 2017).

Aos olhos do RGPD, o ônus da administração de dados pessoais repousa em controladores e processadores de dados singulares que lidam com silos de dados singulares. “A inovação tecnológica que nos trouxe *blockchains* pode, contudo, transformar indivíduos em soberanos de dados que podem copiar, alterar, compartilhar, mover seus dados” (Fink, 2017).

Embora seja inevitável algum grau de transparência em um DLT para permitir que a rede atinja o consenso descentralizado, a transparência é inevitável apenas na camada mais básica do “livro razão” (que na prática é a base de dados em *blockchain*). Camadas adicionais de criptografia e ofuscação podem ser construídas para esconder dados pessoais, de forma elevar o grau de proteção. “Neste contexto, aqueles que são chamados a interpretar e aplicar o RGPD não devem, obviamente, confiar cegamente nos DLTs. É também o papel dos reguladores garantir que essas considerações sejam incorporadas ao software desde o início” (Fink, 2017).

Outro aspecto relevante do direito ao esquecimento, é a possibilidade de perda de outros tipos de dados, por exemplo, no retalho pode ser relevante manter a informação dos produtos e customizações realizadas, isso pode permitir às organizações manterem um *tracking* as tendências de compras. Como no Bitcoin, a técnica de criptografia pode dar resposta a esta questão, “enquanto as chaves públicas de todas as transações, também conhecidas como endereços Bitcoin, são registradas na *blockchain*, tais chaves não são vinculadas à identidade de ninguém. Porém, se a identidade de uma pessoa estivesse associada a uma chave pública, poderíamos vasculhar as transações na *blockchain* e facilmente ver todas as transações associadas a essa chave” (Emanuel, 2014).

A adoção da tecnologia *blockchain*, como repositório de dados, implica a implementação das políticas de segurança assim como o assegurar o controle da rede DLT.

“Numa DLT pública não existe um supervisor, e cada sujeito trabalhando no *blockchain* é o dono de seu (s) nó (s). Nesse caso, de fato, não há um controlador porque o proprietário do nó não pode ser o controlador de si mesmo. Nesta situação, aparentemente, pode parecer que a lei de privacidade e proteção de dados não é aplicável” (Fabiano, 2018).

Apesar de ser uma possível brecha do regulamento, é uma situação de pouca transparência por parte da empresa que toma os dados. “Projetando e configurando uma rede *blockchain*, significa que devem ser criadas políticas de privacidade e segurança aplicáveis a todos os proprietários do nó. Esta solução poderia mitigar a falta da lei onde não é possível aplicá-la ao sistema *blockchain*” (Fabiano, 2018).

Uma questão política deve ser considerada para a discussão da privacidade *by design*. Adotar e implementar totalmente o paradigma de *privacy by design*, é necessário reconhecer a transparência como um atributo importante não só aos dados em si, mas também o *software* que manipula os dados pessoais (código aberto).

Saber o que um sistema faz com nossos dados é a única maneira de permitindo que sujeitos de dados instruídos identifiquem os riscos.

“Por esta razão, escolhemos deliberadamente representar o conceito do consentimento do titular dos dados, de modo que a responsabilidade de fornecer dados pessoais reside, legal e tecnicamente, em suas próprias mãos” (Wirth & Kolain, 2018). Transparência não apenas nos mecanismos e processos de tratamento de dados, mas também transparência do *software* que está a realizar. Sendo um tema técnica, mas de qualquer forma é um modelo de transparência.

Normalmente as redes sociais com maior número de pessoas associadas devem promover iniciativas que podem levar à adoção da tecnologia de *blockchain*.

O *website* Fortune – “*Facebook Is Moving Into Blockchain Technology. Why It Matters*” (Marinova, 2018) – apresenta a iniciativa como resposta aos incidentes com a empresa Cambridge Analytica. O Facebook está aberto para alavancar o *blockchain* para ajudar os usuários a controlar (e talvez até) monetizar seus dados.

6. Conclusões

O RGPD e blockchains compartilham a motivação de capacitar indivíduos e reduzir a assimetria entre eles e as organizações que processam seus dados e suas transações (Ibáñez, O'Hara & Simperl, 2018).

Para Temple-Raston (2012), a Análise Preditiva é uma área de processamento de dados e que lida com a extração de informações, para que sejam utilizados para prever tendências e padrões de comportamento.

Novos e inovadores modelos de negócios que mudam o controle de empresas para indivíduos, sujeitos às condições de privacidade, podem valer a pena ser explorados, descreve Rubinstein (2012). Reforça a importância da AP na atividade econômica, mas já a data da publicação do autor, há uma preocupação com o direito a privacidade e indiretamente a forma com os dados são utilizados pelas empresas.

As empresas necessitam de um “agente-mobilizador” (Carvalho, 2014) que possa tirar amostras do fluxo de mensagens e procurar tendências e identificar problemas que podem afetar a atividade comercial. Mas a necessidade de transparência e que garanta os direitos individuais são imprescindíveis, quer seja do ponto de vista da regulamentação e como ferramenta que permita melhorar o nível de confiança na relação cliente-fornecedor.

O poder de processamento de dados em tecnologias de Big Data e com técnicas de análise preditiva, vem transformando o mundo. Os exemplos são muitos, sejam pela positiva ou negativa. A influência das redes sociais alterou o funcionamento de diversos mercados e o tema da “transformação digital” passou a estar nos planos estratégicos das empresas e governos. A regulamentação vem em resposta a estes novos tempos, para atender a necessidade individual da privacidade e não violação de direitos.

A tecnologia blockchain possui características interessantes e que aparentam responder a muitas das questões de privacidade e da regulamentação. As características de funcionamento distribuído e de segurança (estando há mais de uma década em funcionamento a suportar o Bitcoin e até a data não teve a segurança quebrada). A blockchain, por este histórico com a Bitcoin, apresenta uma estrutura já testada e com certa maturidade, o que reduz em muitos custos com novos desenvolvimentos. No entanto a fase de adaptação será sempre necessária, mas não é o mesmo do que partir de um ponto zero.

A pesquisa procurou elucidar as seguintes questões:

Q1 *A análise preditiva é relevante para as organizações?*

Os resultados da análise desta investigação vão de encontro com os resultados do trabalho de Pereira (2014, quando demonstra a importância do processo de descoberta de conhecimento a partir de bases de dados, a passar por uma série de etapas, desde a coleta, a mineração dos dados, a consolidação e extração dos padrões e regras, e por fim a agregação de valor que possibilita uma melhor tomada de decisão).

Esta avaliação permitiu compreender que os recursos de tecnologia da informação, nomeadamente as plataformas de *big data* e análise preditiva são indispensáveis ao suporte ao processo de decisão. Principalmente em instituições privadas e órgãos públicos que atuam de forma cada vez mais dinâmica, e que requerem monitoramento constante, pois com a velocidade que as comunicações e interações acontecem assim o exige. E para se realizar tal feito é indispensável à correta aplicação destes recursos, pois só assim é possível garantir um eficiente e eficaz sistema de apoio à decisão.

Se pode verificar que a utilização adequada de sistemas e técnicas de predição são capazes de fornecer informações confiáveis, úteis e tempestivas ao processo decisório, tornando-se em um instrumento com alto grau de confiabilidade e com benefícios mensuráveis para as empresas e organizações.

Q2 *Qual a importância da RGPD nas organizações?*

As novas tecnologias não mudam apenas como aplicamos as regulamentações existentes a novos fatos, mas também pode desestabilizar profundamente as bases sobre as quais a regulamentação existente se baseia. Aos olhos do RGPD, o ônus da administração de dados pessoais repousa em controladores e processadores de dados singulares que lidam com silos de dados, ou seja, empresas/organizações, como afirma Fink (2017, em um trabalho de análise ao regulamento). O RGPD permite as organizações, sobre tudo, tornarem os processos mais transparentes, e é uma oportunidade para elevar o nível de confiança. Sendo a confiança um requisito fundamental para qualquer tipo de relação.

Para além do aspecto regulatório, a garantia de não violação dos direitos individuais, permitem uma redução de tensões no mercado. “Todas as pessoas têm direito ao respeito

pela sua vida privada e familiar, pelo seu domicílio e pelas suas comunicações.”, Artigo 7º do RGPD.

Q3 A tecnologia blockchain, pode assegurar a realização de tratamento de dados face ao RGPD?

A regulamentação e a tecnologia de blockchain, partilham a mesma motivação de capacitar indivíduos e reduzir a assimetria entre pessoas e as organizações que processam seus dados e suas transações. A blockchain parece acomodar o RGPD, pois seus membros podem ser designados como controladores conjuntos de dados, assumindo a responsabilidade pela proteção de dados, sendo o mesmo resultado dos estudos do Ibáñez, O’Hara & Simperl (2018, “*General Data Protection Regulation and the Blockchains*”).

A blockchain, tem entre outros o lado positivo de não haver a necessidade de questionar a conformidade com tarefas estatutárias (nacionais), e é possível apagar ou alterar dados pessoais. A propriedade das informações mais sensíveis e pessoais (caracterizadas no regulamento) estariam salvaguardadas a pessoa, sendo esta responsável pela sua manutenção, no entanto, permite as empresas utilizarem os rastros, ou transações efectuadas pelas pessoas. O que permite a continuidade da capacidade de análise de dados e da AP como ferramenta de apoio a decisão.

6.1 Limitações e sugestões para investigação

A identificação de fatores críticos de sucesso é relevante para empresas que desejem tirar proveito da Transformação Digital que o mercado está a presenciar a data.

Os estudos de caso apoiam futuras pesquisas que visem a elaboração de um modelo a partir das variáveis identificadas e a generalização dos conhecimentos obtidos. Contudo, é importante a realização de estudos com abordagem quantitativa para verificar a validade e aplicabilidade de blockchain como mecanismo de armazenamento de dados e de conformidade com a regulamentação.

7. Bibliografia

- Angeloni, M. T. (2003). Elementos interveripientes na tomada de decisão. *Ci. Inf., Brasília*, 32(1998), 17–22.
- Braga, A. M. (2016). Segurança e Desenvolvimento de Software . Introdução. *Cpqd*, 34. Retrieved from www.cpqd.com.br
- Brynjolfsson, E., & McAfee, A. (2012). Big Data : The Management Revolution. *Harvard Business Review*, (October).
- Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of ‘big data.’ Retrieved from <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/seizing-the-potential-of-big-data>
- Carvalho, R. M. (2014). *A Força das Coisas: Diário de um futuro lusófono*. BNOMICS.
- Chicarino, V. R. L., Ferreira Jesus, E., Albuquerque, C. V. N., & Rocha, A. A. (2017). Uso de Blockchain para Privacidade e Segurança em Internet das Coisas, (November). Retrieved from https://www.researchgate.net/profile/Vanessa_Rocha_Leandro_Chicarino/publication/321966650_Uso_de_Blockchain_para_Privacidade_e_Seguranca_em_Internet_das_Coisas/links/5a3b92aaaca272774f9baf5a/Uso-de-Blockchain-para-Privacidade-e-Seguranca-em-Internet-das-
- Ciccarino, I., & Araki, M. E. (2017). Blockchain como um fator demudança na competição e no arranjo econômico de oportunidades. *Xx Semead*, 1(Fall), 0–14.
- Colombo, C., & Neto, E. F. (2017). Revista de Direito, Governança e Novas Tecnologias. *Revista de Direito, Governança e Novas Tecnologias*, 59–80.
- Emanuel, D. (2014). Aspectos de segurança na rede Bitcoin, 221–249. Retrieved from http://www.cdn.ueg.br/source/observatorio_inhumas/conteudoN/3322/CAP_8_CR_ASPECTOS_DE_SEGURANCA_NA_REDE_BITCOIN.pdf
- Europeia, S. das P. da U. (2018a). Reforma em matéria de proteção de dados na UE : proteção de dados reforçada para os cidadãos europeus. <https://doi.org/10.2838/874736>
- Europeia, S. das P. da U. (2018b). Reforma em matéria de proteção de dados na UE: regras melhores para as empresas europeias.
- Fabiano, N. (2018). Blockchain and Data Protection: the value of personal data - NicFab. Retrieved June 17, 2018, from <https://www.nicfab.it/blockchain-data-protection/>
- Faceli, K., Lorena, A. C., Gama, J., & Carvalho, A. C. P. L. F. (2011). *Inteligência artificial : uma abordagem de aprendizado de máquina*. Rio de Janeiro: Grupo Gen - LTC.
- Ferreira, L. M. (2018). Facebook enfrenta novo teste. Lucros resistem ao escândalo Cambridge Analytica? Retrieved April 26, 2018, from <http://www.jornaleconomico.sapo.pt/noticias/facebook-enfrenta-novo-teste-lucros-resistem-ao-escandalo-cambridge-analytica-297700>
- Fink, M. (2017). Blockchains and Data Protection in the European Union. *Max Planck Institute for Innovation and Competition Research Paper No. 18-01*, 18(1), 1–30. <https://doi.org/10.2139/ssrn.3080322>

- Franks, B., & Davenport, T. (2012). *Taming the big data tidal wave*. New Jersey: John Wiley & Sons, Inc. Hoboken.
- Geelkerken, F. W. J. Van, & Konings, K. (2017). Using Blockchain to strengthen the rights granted through the GDPR. *INTERNATIONAL YOUTH SCIENCE FORUM "LITTERIS ET ARTIBUS"*, 23–25 NOVEMBER 2017, LVIV, UKRAINE, (November 2017), 23–25.
- Gil, A. C. (1991). *Como elaborar projetos de pesquisa* (3ª). São Paulo: Atlas.
- Gonçalves, A., Pereira, D., Pinheiro, E., & Aguiar, P. (2008). *Inteligência Artificial & Data Mining: Projeto de Sistema de Apoio à Decisão*. Rio de Janeiro.
- Guazzelli, A. (2012). O Que é a Análise Preditiva? Retrieved May 15, 2018, from <https://www.ibm.com/developerworks/br/industry/library/ba-predictive-analytics1/>
- Hilbert, M., & Lopez, P. (2011). *A Capacidade Tecnológica do Mundo para Armazenar, Comunicar as Informações*. Editora Science 332.
- IBM. (2013). Saiba o que é o Big Data e os desafios que as empresas enfrentam. Retrieved from http://www.ibm.com/midmarket/br/pt/infografico_bigdata.html
- IDC. (2011). The 2011 Digital Universe Study: extracting value from chaos. Retrieved from <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
- Katori, F. Y. (2017). Impactos Das Fintechs E Do Blockchain No Sistema Financeiro: Uma Análise Crítico-Reflexiva.
- Kauark, F. da S., Manhães, F. C., & Medeiros, C. H. (2010). Metodologia da Pesquisa: um guia prático. *Via Litterarum*, 88.
- Lakatos, E. M., & Marconi, M. de A. (1991). *Fundamentos de Metodologia Científica* (3rd ed.). São Paulo: Atlas.
- Lindstrom, M. (2017). *Small Data*. (HarperCollins Brasil, Ed.) (1ª). HarperCollins Brasil. Retrieved from <https://www.amazon.com.br/Small-Data-Martin-Lindstrom/dp/8569809719>
- Machado, M. P. (2018). “O Brexit nunca teria acontecido sem a Cambridge Analytica”, revela ex-funcionário. Retrieved April 25, 2018, from <https://observador.pt/2018/03/26/o-brexite-nunca-teria-acontecido-sem-a-cambridge-analytica-revela-ex-funcionario/amp/>
- Marinova, P. (2018). Facebook Is Moving Into Blockchain Technology. Why It Matters | Fortune. Retrieved June 10, 2018, from <http://fortune.com/2018/05/09/facebook-blockchain-team/>
- Marr, B. (2018). Here’s Why Data Is Not The New Oil. Retrieved June 9, 2018, from <https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#4c87ed0d3aa9>
- Martins, P. (2018). *Introdução à Blockchain* (1ª Edição). Lisboa: FCA.
- Mota, J. dos R., & Siqueira, L. N. P. (2016). COMINE – XII Congresso Mineiro de Empreendedorismo, 1–15.
- Notícias, J. de. (2017). Uber pagou a “hackers” para apagarem dados de clientes roubados. Retrieved April 26, 2018, from <https://www.jn.pt/mundo/interior/uber-pagou-100000-dolares-para-hackers-apagarem-dados-de-utilizadores-pirateados-8934911.html>

- Novo, R., & Neves, J. M. S. (2013). Inovação na inteligência analítica por meio do Big Data : características de diferenciação da abordagem tradicional, 32–44.
- Palatinick, R. (2017). Blockchain: Collaborate to innovate. Retrieved May 2, 2018, from <https://www.euromoney.com/article/b12khp84ytkk2n/blockchain-collaborate-to-innovate>
- Paulo, L., & Cesar, D. (2009). O Pensamento Científico - Abordagem Da Pesquisa No Estudo Interpretativo. *Revista de Educação, XII*(13), 99–107.
- Pembernton, C. (2018). Pay Attention to These 6 Marketing Technologies in 2018. Retrieved May 2, 2018, from <https://www.gartner.com/smarterwithgartner/pay-attention-to-these-6-marketing-technologies-in-2018/>
- Pilkington, M. (2015, September 18). Blockchain Technology: Principles and Applications. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2662660
- Rodrigues, C. K. da S. (2017). A Simple Analysis of Efficiency and Security of the Blockchain Technology . *Computing and System Journal, 7*(2), 147–162. Retrieved from <http://www.revistas.unifacs.br/index.php/rsc>
- Ross, A. (2016). *As Indústrias do Futuro*. Actual.
- Rotella, P. (2012). Is Data The New Oil? Retrieved June 9, 2018, from <https://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/#1df803187db3>
- Rubinstein, I. S. (2013). Big Data : The End of Privacy or a New Beginning ? *International Data Privacy Law, (12)*.
- Saldanha, N. (2018). *Novo Regulamento Geral de Proteção de Dados* (1ª Edição). Lisboa: FCA.
- Schwab, K. (2016). *A Quarta Revolução Industrial* (World Econ). Cologny, Suíça: Levoir.
- Siegel, E. (2016). *Predictive Analytics*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Silveira, M., Marcolin, C. B., & Freitas, H. M. R. (2015). Uso Corporativo do Big Data : Uma Revisão de Literatura. *Revista de Gestão e Projetos - GeP, 6*(3), 44–59. <https://doi.org/10.5585/gep.v6i3.369>
- Smolan, R., & Erwitte, J. (2013). The Human Face of Big Data. Retrieved May 31, 2018, from https://www.pbslearningmedia.org/resource/bigdata_stem_flu_patterns/the-human-face-of-big-data-google-predicts-the-flu/#.WxAbl1MvzOQ
- Tam Vo, H., Kundu, A., & Mohania, M. (2018). Research Directions in Blockchain Data Management and Analytics, 445–448. <https://doi.org/10.5441/002/edbt.2018.43>
- Taurion, C. (2013). *Big Data*. Rio de Janeiro: Brasport.
- Tessarolo, P. H., & Magalhães, W. B. (2014). a Era Do Big Data No Conteúdo Digital : Os Dados Estruturados E Não Estruturados.
- Ulrich, F. (2014). *Bitcoin: a moeda na era digital*. São Paulo: Instituto Ludwig von Mises Brasil.
- Vasconcelos, J. B., & Barão, A. (2017). *Ciência dos Dados nas Organizações* (1st ed.). Lisboa: Editora de Informática, Lda.

Volpato, T., Rufino, R. R., & Dias, J. W. (2013). Big Data – Transformando Dados Em Decisões.

Wirth, C., & Kolain, M. (2018). Privacy by BlockChain Design: A Blockchain-enabled GDPR-compliant Approach for Handling Personal Data.
https://doi.org/10.18420/blockchain2018_03

Yin, R. K. (1994). *Case study research : design and methods*. Sage Publications.
Retrieved from
https://books.google.pt/books/about/Case_study_research.html?id=AvYOAQAAMAAJ&redir_esc=y

Zago, M. G. (2018). 50+ Examples of How Blockchains are Taking Over the World.
Retrieved June 2, 2018, from <https://medium.com/@matteozago/50-examples-of-how-blockchains-are-taking-over-the-world-4276bf488a4b>