

**CONTROLE DE CUSTOS ASSISTENCIAIS NA SAÚDE SUPLEMENTAR UTILIZANDO BIG DATA E ANALYTICS PARA PREVER COMPORTAMENTOS E ANTECIPAR CUIDADOS AOS BENEFICIÁRIOS**

**CONTROL OF SUPPLEMENTARY HEALTHCARE COSTS USING BIG DATA AND ANALYTICS TO PREDICT BEHAVIOR AND ANTICIPATE CARE TO BENEFICIARIES**

**CONTROL DE LOS COSTOS SUPLEMENTARIOS DE SALUD MEDIANTE EL USO DE BIG DATA Y ANALYTICS PARA PREDECIR EL COMPORTAMIENTO Y ANTICIPAR EL CUIDADO A LOS BENEFICIARIOS**

**Edmar Oliveira Guedes Júnior**  
Universidade Federal do Ceará  
contato@edmarguedes.com.br

**Francisco Alberto de Oliveira Júnior**  
Universidade Federal do Ceará  
alberto.oliveira@unimedfortaleza.com.br

**Érico Veras Marques**  
Universidade Federal do Ceará  
ericovmarques@gmail.com

**Jocildo Figueiredo Correia Neto**  
Universidade Federal do Ceará  
jocildo@ufc.br



## RESUMO

O controle dos custos assistenciais é uma preocupação central das operadoras de saúde no Brasil, especialmente após a Lei 9656/98 e a criação da Agência Nacional de Saúde. Foram incluídas diversas obrigações no atendimento aos usuários, exigindo das operadoras adequações administrativa e financeira. Diante disso, é importante adotar estratégias para o controle de custos, para buscar viabilidade financeira no médio e no longo prazo. As ferramentas em tecnologia da informação trazem novas perspectivas de atuação no controle de desembolsos assistenciais, já que viabilizam estratégias para otimizar processos e antecipar cuidados aos beneficiários dos planos de saúde. Este estudo apresenta o resultado obtido com a aplicação das ferramentas de *Big Data* e *Data Analytics* na redução de notificações de intermediação preliminar (NIPs) e liminares judiciais em uma grande operadora de saúde. Utilizou-se análise de conteúdo sobre documentos de uma grande operadora de saúde. Observou-se que essas ferramentas conseguem prever o comportamento de beneficiários no tocante à abertura de NIPs e liminares, contribuindo para a melhoria da qualidade do atendimento aos clientes e à redução de gastos.

**Palavras-chave:** custos assistenciais; operadoras de saúde; big data.

## ABSTRACT

The control of health care costs is a central concern of healthcare operators in Brazil, especially after the Law 9656/98 and the creation of the National Health Agency. Several obligations were included in services to users, requiring operators' administrative and financial adjustments. Therefore, it is important to adopt cost control strategies in order to seek financial viability in the medium and long term. The information technology tools bring new perspectives of action in the assistance disbursements control, since they enable strategies to optimize processes and anticipate care for the beneficiaries of health plans. This study presents the results obtained with the application of Big Data and Data Analytics tools in the reduction of preliminary intermediation notifications (PINs) and court injunctions in a large healthcare operator. Content analysis on documents from a big healthcare operator was used. It was observed that these tools can predict the behavior of beneficiaries with regard to the opening of PINs and injunctions, contributing to improving the quality of customer service and reducing expenses.

**Keywords:** assistance costs; health care companies; big data.

## RESUMEN

El control de los costos de la atención médica es una preocupación central de los operadores de atención médica en Brasil, especialmente después de la Ley 9656/98 y la creación de la Agencia Nacional de Salud. Se incluyeron varias obligaciones en el servicio a los usuarios, que requieren reajustes administrativos y financieros de los operadores. Por lo tanto, es importante adoptar estrategias de control de costos para buscar la viabilidad financiera a mediano y largo plazo. Las herramientas en tecnología de la información permiten nuevas perspectivas de acción en el control de los desembolsos de asistencia, ya que permiten estrategias para optimizar procesos y anticipar la atención a los beneficiarios de los planes de salud. Este estudio presenta los resultados obtenidos con la aplicación de herramientas de Big Data y Data Analytics en la reducción de notificaciones preliminares de intermediación (NPIs) y órdenes judiciales en un gran operador de atención médica. Se utilizó el análisis de contenido de documentos de un grande operador de atención médica. Se observó que estas herramientas son capaces de predecir el comportamiento de los beneficiarios con respecto a la apertura de NPIs y medidas cautelares, contribuyendo a la mejora de la calidad del servicio al cliente y la reducción de gastos.

**Palabras clave:** costos de asistencia; empresas de servicios de salud; big data.

## INTRODUÇÃO

O impacto dos custos assistenciais vem sendo uma das maiores preocupações das operadoras de planos de saúde no Brasil, em especial após a criação da Agência Nacional de Saúde (ANS) em janeiro de 2000. Sua visão reguladora do mercado gerou uma rede de proteção e segurança aos beneficiários, porém aumentou as exigências em cima de diretrizes e normas operacionais junto às prestadoras (MALTA *et al.*,

2004). Segundo Leal e Matos (2009), a relevância da evolução dos custos em saúde é um fenômeno mundial, num contexto de tendência de participação crescente no PIB das nações.

Malta *et al.* (2004) afirmam que a expansão do mercado de Saúde Suplementar nas últimas décadas foi significativa, estimando-se, segundo os dados da PNAD/98, em 38,7 milhões, o número de brasileiros cobertos por, pelo menos, um plano de saúde, correspondendo a cerca de 24,5% da população do País. Em dezembro de 2019 o setor contabilizou 47.039.728 beneficiários ou 23,2% da população nacional em planos de assistência médica e 26.024.494 ou 13,3% da população em planos exclusivamente odontológicos (ANS, 2020b).

A incorporação de procedimentos antes não cobertos e a constante inovação tecnológica em novos equipamentos e tratamentos, sem a devida equivalência nas contraprestações dos clientes, exigiram das organizações de saúde suplementar uma readequação de suas estruturas internas, de modo que a sua capacidade financeira permita atender às exigências e preserve seu equilíbrio (ALBUQUERQUE, 2008).

Ocké-Reis, Andreazzi e Silveira (2006) destacam que os custos associados à incorporação de equipamentos contribuem para essa tendência de crescimento dos gastos. O modelo assistencial acentua tal incorporação estimulando, por exemplo, o uso indiscriminado de exames como a tomografia computadorizada e a ressonância magnética, em particular porque não há um controle rigoroso das suas indicações e nem mesmo da incorporação e utilização das novas tecnologias e tratamentos no sistema de saúde suplementar.

A peculiaridade do setor saúde em relação às outras empresas de mercado exigiu a readequação de processos internos que fizeram muitas operadoras passarem por dificuldades financeiras. Muitas não resistiram e sucumbiram, sendo incorporadas a outras operadoras ou foram simplesmente extintas. Segundo dados da ANS, havia 1.135 operadoras de saúde no Brasil em 2008. Atualmente, apenas 740 estão funcionando, sendo que algumas em situação de dificuldade (ANS, 2020a; ANS, 2020b).

Oliveira (2011) destaca que, apesar da legislação criteriosa e das regras impostas pela ANS, observam-se ainda comportamentos reprováveis por parte de algumas operadoras na limitação da cobertura assistencial ou na burocracia excessiva. Por outro lado, os usuários passaram a ter mais consciência de seus direitos e, portanto, o volume de reclamações aumentou, inclusive além do estabelecido na legislação regulamentadora e nos contratos, como alguns tipos de cirurgias consideradas estéticas, tratamentos para fertilização e próteses não ligadas aos atos cirúrgicos. A consequência dessas duas posições antagônicas acabou por gerar grande número de demandas judiciais em todo o País. Com o advento do Código de Defesa do Consumidor, as decisões judiciais passaram a considerar o usuário hipossuficiente, gerando um volume maior de resultados em favor deste, gerando insatisfação e prejuízo às operadoras. É preciso reconhecer que as alterações da legislação, provocadas pela ANS, tiveram o objetivo de reduzir as demandas judiciais e tornar as regras mais claras para a concessão de benefícios de saúde.

O crescente e constante aumento do volume de processos judiciais levou o Conselho Nacional de Justiça (CNJ) a encaminhar uma recomendação para os tribunais e magistrados acerca do melhor aparelhamento e profissionalização destes, no intuito de tomarem decisões mais adequadas às questões tanto relacionadas com a saúde privada quanto com a pública. O problema é que, em razão dos custos ilimitados, as questões são de difícil solução (OLIVEIRA, 2011).

A justificativa para a realização desse estudo sobre o tema do uso de novas tecnologias como o *big data* e *data analytics*, a fim de reduzir custos e despesas assistenciais, lastreia-se na importância e dos seus reflexos sobre a assistência em saúde da população. Ademais, percebe-se a escassa literatura sobre o assunto, o que abre oportunidades acadêmica e profissional de exploração científica.

A questão de pesquisa basilar é: o uso das ferramentas de *big data* e *analytics* podem contribuir com a redução de notificações de intermediação preliminar (NIPs) e liminares judiciais contra a Unimed Fortaleza?

Diante desse contexto, o presente estudo tem como objetivo analisar o resultado obtido a partir da aplicação das ferramentas tecnológicas de *Big Data* e *Data Analytics* como modelo preditivo do comportamento de clientes, a fim de reduzir o número de NIPs e liminares judiciais em uma operadora de saúde de grande porte do sistema Unimed.

Como objetivos específicos, têm-se: apresentar os passos para a implantação das ferramentas e apresentar os resultados alcançados com o uso das ferramentas na redução de NIPs e liminares judiciais contra a operadora de saúde. Para isso, lança mão de um estudo qualitativo com abordagem descritiva, a partir de dados colhidos da oitava maior operadora do sistema Unimed.

O artigo dividiu-se em partes, sendo a primeira esta introdução. Na segunda parte, discorre-se sobre custos assistenciais. Na terceira parte, busca-se a familiarização com o sistema de saúde suplementar no Brasil, contemplando as peculiaridades do marco regulatório exercido pela ANS. Na quarta parte, apresentam-se as ferramentas de *Big Data* e *Data Analytics*. Na quinta parte, elencam-se os aspectos metodológicos, na sexta, a exposição dos resultados e discussão e na sétima parte, as considerações finais.

## REFERENCIAL TEÓRICO

O referencial teórico necessário para embasar este estudo é composto inicialmente pelos custos assistenciais, seguido pelo sistema de saúde suplementar no Brasil e finalizado com *big data* e *analytics*. Estes tópicos são apresentados a seguir, nesta ordem.

### Custos Assistenciais

Grande atenção é dada em relação ao tema de custos assistenciais por parte das operadoras de planos de saúde, decorrente da necessidade de aperfeiçoar e adequar seus recursos para que se possa proporcionar o atendimento à sua carteira de clientes com todos os itens exigidos pela Agência Nacional de Saúde (ANS). O não cumprimento dessas adequações resulta em severas e indesejáveis penalidades (MALTA *et al.*, 2004).

O monitoramento contínuo dos custos se torna essencial, devendo-se lançar mão da adoção de várias ferramentas gerenciais. Segundo Taveira (1999), o gestor precisa dispor de informações que permitam a ele responder, a qualquer instante, sobre, minimamente, dois aspectos básicos do seu processo de condução. O primeiro é a eficiência com a qual os serviços vêm sendo operacionalizados em seu estabelecimento. A segunda, não menos importante, é o nível de eficácia atingido por eles.

Segundo a ANS (BRASIL, 2013), denomina-se custo assistencial o desembolso resultante de toda e qualquer utilização das coberturas contratadas junto aos planos de saúde suplementar por parte do beneficiário.

Na Resolução normativa número 435 de 23 de novembro de 2018, a ANS descreve os itens que compõem os custos assistenciais para operadoras de saúde, os quais são apresentados no Quadro I.

**Quadro I** - Quadro resumo com os itens da RN 435 da ANS.

Item	Especificação
Consultas médicas	Atendimento prestado por médico em regime ambulatorial
Outros atendimentos ambulatoriais	Atendimento realizado em regime ambulatorial realizado por profissionais de nível superior não médicos.
Exames	Procedimento de auxílio diagnóstico em regime ambulatorial.
Terapias	Atendimento utilizando métodos de tratamento, em regime ambulatorial.
Internações	Internação em leito hospitalar (enfermaria, quarto ou unidades de curta permanência, terapia intensiva ou semi-intensiva).
Demais despesas médico-hospitalares	Despesas assistenciais médico hospitalares acessórias.
Procedimentos odontológicos	Atendimento com fins de diagnóstico e orientação terapêutica em saúde bucal, em regime ambulatorial.

Fonte: Elaborado pelos autores (2020) (ANS, 2019c).

Diversos indicadores de acompanhamento contábil são utilizados para controle dos custos frente às receitas. Ressalte-se, dentre estes, o índice de Variação de Custo Médico-Hospitalar (VCMH), que expressa a variação do custo das operadoras de planos de saúde comparado entre dois períodos consecutivos de 12 meses. A variação de custos médico-hospitalares é composta principalmente pela variação de gastos em itens como o de internações, que possui o maior peso no índice, compondo 61% dos custos. A VCMH/IESS é composta ainda pelos seguintes procedimentos: exames complementares (13%), consultas (9%), terapias (11%) e outros serviços ambulatoriais (6%). O referido índice captura as oscilações decorrentes da frequência de utilização e do aumento de preços dos procedimentos médicos (IESS, 2019).

O crescente aumento das despesas por conta de inovações tecnológicas, com novas aparelhagens, disponibilizadas para o auxílio ao diagnóstico médico, gera por vezes, procedimentos desnecessários e de alto custo. Para os planos privados de assistência, os exames mais sofisticados impactam significativamente em seu resultado. A chamada inflação tecnológica para diagnóstico e tratamento das doenças, somam-se ainda como ofensores dos custos: o envelhecimento populacional, os processos judiciais lançados contra os prestadores e a própria característica da vida moderna, com seus hábitos e práticas nocivas: stress, violência e sedentarismo, que só pioram essa situação (MIRANDA, 2004; DUARTE *et al.*, 2017).

### Sistema de Saúde Suplementar no Brasil

Essa seção foi estruturada visando familiarizar o leitor sobre a evolução dos serviços de saúde suplementar no Brasil, que passou por profundas transformações ao longo das décadas. A seguir, define o conceito de saúde suplementar detalhando as diversas modalidades de operadoras de saúde. Por fim, narra-se o surgimento da regulação do setor, a partir da implantação da lei 9656 de 1998, seguida da criação da Agência Nacional de Saúde (ANS) e seus efeitos junto às operadoras de saúde.

A saúde suplementar no Brasil passou por profundas transformações ao longo do tempo. Nos anos 40, surgem os Institutos de Aposentadorias e Pensões (IAPs), que pertenciam a diversas categorias e representavam os trabalhadores urbanos, comprando as prestações de serviços de saúde, além das Caixas de Assistência de inúmeras instituições (PIETROBON; PRADO; CAETANO, 2008).

Nos anos 50, empresas estatais e multinacionais começaram a oferecer os serviços de saúde ou utilizavam-se do serviço de reembolso para atender aos seus funcionários. Na década de 60, os serviços de atenção à saúde convivem de maneira paralela, atendendo grande parte do contingente de trabalhadores, e funda-se o INPS, que atendia de maneira global à população. Na década de 70, com a fundação do INAMPS, a compra de serviços médicos junto a prestadores privados aumentou ainda mais a disponibilidade de capital no setor de saúde. A década de 80 foi marcada pela precarização do atendimento em saúde pública, gerando um aumento na procura por serviços privados de atenção à saúde. A deficiência do Estado em fornecer atendimento adequado à população e a maior procura por atendimento privado gerou uma capitalização das operadoras de planos de saúde, sem o correspondente atendimento adequado ao público consumidor. Apesar da implantação do Código de Defesa do Consumidor, que trouxe a estrutura jurídica a nortear as relações de consumo e, notadamente, a atuação nos Procons estaduais e municipais, ainda havia conflitos entre operadoras e clientes no fornecimento de serviços que não estavam sendo prestados. Nesse contexto, tornavam-se necessários novos mecanismos reguladores dessa delicada relação envolvendo a saúde da população naquele momento (RONCALLI, 2003).

Define-se como suplementar o setor de planos de saúde no Brasil, pelo fato de se pagar um seguro privado para se ter acesso à assistência médica, mesmo mantendo a contribuição compulsória para a seguridade social, que inclui o direito ao acesso ao serviço público. Além disso, seu caráter supõe a existência e a limitação do sistema de saúde público, no qual o sistema privado complementaria a cobertura de determinados serviços (BAHIA, 2001a).

Segundo a ANS (2020a), as operadoras se dividem nas modalidades indicadas no Quadro 2.

**Quadro 2** – Quadro resumo com os tipos de operadoras de saúde.

Item	Especificação
Administradora de planos	Administram planos de assistência à saúde, mas não possuem beneficiários.
Administradora de benefícios	Propõe a contratação de plano coletivo na condição de estipulante.
Autogestão	Opera serviços de assistência à saúde através de gestão própria.
Cooperativa médica	Operadora formada por médicos que se constitui na forma de associação de pessoas sem fins lucrativos.
Cooperativa odontológica	Operadora formada por odontólogos que se constitui em associação de pessoas sem fins lucrativos.
Filantropia	Operadora em entidade sem fins lucrativos.
Medicina de grupo	Operadora que se constitui em sociedade que comercializa ou opera planos privados de saúde, excetuando-se as classificadas nas modalidades administradora, cooperativa médica, autogestão, filantropia e seguradora especializada em saúde.
Odontologia de grupo	Operadora que se constitui em sociedade que comercializa ou opera planos odontológicos.
Seguradora especializada em saúde	Constituída em sociedade seguradora com fins lucrativos que comercializa seguros de saúde.

Fonte: Elaborado pelos autores (2020).

A regulação do setor de saúde suplementar surgiu em 1998, com a publicação da Lei n. 9656/98 (BRASIL, 1998) que buscava corrigir a falta de mecanismos de controle do governo federal sobre este setor. Embora apresentando avanços na proteção aos usuários de planos de saúde, a Lei 9656/98 não sanou os principais conflitos entre as operadoras e seus clientes (SALAZAR; RODRIGUES; NUNES JÚNIOR, 2005).

Com isso, em janeiro de 2000, é criada a Agência Nacional de Saúde Suplementar (ANS), por meio da Lei 9.961/2000. Trata-se de uma autarquia sob regime especial, vinculada ao Ministério da Saúde, responsável pela regulação, normatização, controle e fiscalização das atividades que garantam a assistência suplementar à saúde (BRASIL, 2000).

Segundo Bahia (2001b), existem divergências quando se discute qual é o objeto e a intensidade dessa regulação. Para alguns, a regulamentação visa corrigir/atenuar as falhas do mercado com relação à assimetria de informações entre clientes, operadoras e provedores de serviços. A regulação deveria então atuar minimizando a seleção de riscos, por parte das empresas de planos, que preferem propiciar cobertura aos riscos “lucrativos” e por parte de clientes, que tendem a adquirir seguros/planos, em razão de já apresentarem alguma manifestação do problema de saúde preexistente.

O modelo de regulação do setor de saúde suplementar brasileiro é diferenciado em relação à experiência internacional. Na maioria dos países, a regulação é feita a partir da atividade econômica em si, atuando sobre as empresas do setor e garantindo suas condições de solvência e a competitividade do mercado. No Brasil, optou-se por regular fortemente o produto oferecido, ou seja, a assistência à saúde, com medidas inovadoras, como a proibição da seleção de risco e do rompimento unilateral dos contratos (BAHIA; VIANA, 2002).

Um dos mecanismos utilizados na regulação é a Notificação de Intermediação Preliminar (NIP), que, segundo o Art 5º, RN 388/2015, é o instrumento que visa a solução de conflitos entre beneficiários e operadoras de planos privados de assistência à saúde, constituindo uma fase pré-processual. Pode-se classificá-la em dois tipos: NIP assistencial, que se refere a toda e qualquer restrição de acesso à cobertura assistencial (assistência ao serviço de saúde), e NIP não assistencial, que se refere a outros temas que não a cobertura assistencial, no qual o beneficiário seja diretamente afetado pela conduta, e a situação seja passível de intermediação, tais como: reajuste, questões contratuais, rede de atendimento, questões referentes a dependentes, reembolso e tempo de carência.

Essa considerável regulamentação do setor provocou forte impacto financeiro nas empresas de saúde suplementar. Segundo Sousa e Salvalaio (2010), além do aumento nas exigências de cobertura de atendimento, a obrigatoriedade da constituição de reservas financeiras visando garantir a continuidade do

atendimento, causou a redução do número de operadoras. De 2009 a 2019, 491 operadoras com beneficiários encerraram suas atividades, apesar de um mercado potencialmente em expansão (ANS, 2020a).

### **Big Data e Analytics**

O acesso à informação e a forma como se trabalha com ela tem se tornado um diferencial competitivo no mundo dos negócios. A proposta da utilização de grande quantidade de informações remonta à década de cinquenta (CHEN; CHIANG; STOREY, 2012). O volume crescente das informações fora das empresas é evidenciado nas redes sociais, sensores em produtos, registros de tráfego de internet, dentre outros. Este grande volume de dados é a característica principal do *Big Data*. Outras características importantes são variedade, velocidade, veracidade e valor (DAVENPORT, 2014; DEMCHENKO *et al.*, 2013).

Autores como Davenport (2014) e Chow-White e Green (2013) retratam o uso do *Big Data* como ferramenta emergente para a tomada de decisão corporativa. O termo *Big Data* surgiu em meados de 2010 para conceituar uma ferramenta tecnológica capaz de trabalhar com grande quantidade de dados, de diferentes fontes e formatos (CHEN; ZHANG, 2014).

Manyika *et al.* (2011) definem *Big Data* como grupo de dados cujo tamanho está além da habilidade de captura, armazenagem e análise de um típico software de banco de dados, e segundo os mesmos autores, a definição de grande é propositadamente subjetiva, já que as características dependem da habilidade e capacidade de cada setor em trabalhar com as informações.

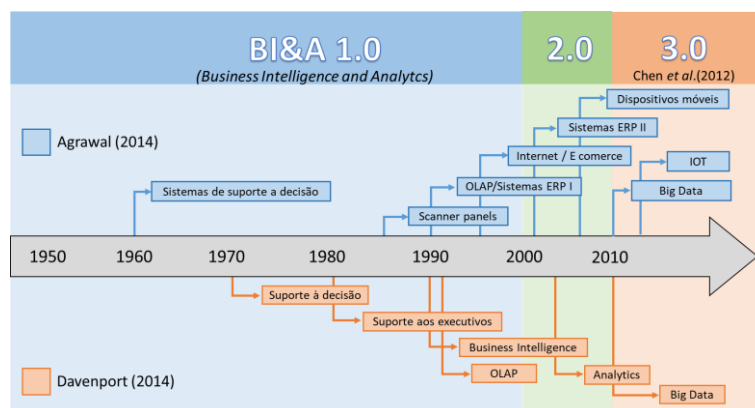
O grande volume de dados é a principal característica do *Big Data*. De acordo com Davenport (2014), somente no ano de 2012, foram gerados no mundo cerca de 2,8 trilhões de gigabytes em dados. Além do volume de dados, outras características também são relevantes, tais como a grande variedade e a alta velocidade da utilização dos dados, (MANYIKA *et al.*, 2011; MCAFEE; BRYNJOLFSSON, 2012). As fontes de dados são diversas e provenientes de dispositivos móveis, celulares, sensores, redes sociais, internet das coisas, dados da própria empresa, imagens, sistemas de monitoramento, registros de software, tuítes, comentários no Facebook, vídeos no YouTube etc.

Demchenko *et al.* (2013) citam outras duas características importantes: a veracidade e o valor. Para os autores, a veracidade trata da necessidade de que os dados coletados tenham origem comprovada e reputação. O valor se traduz em dados que tenham significado para a tomada de decisão na organização.

Para analisar esta infinidade de dados, as empresas terão de desenvolver processos contínuos, a fim de interpretá-los adequadamente (GANDOMI; HAIDER, 2015). Vale ressaltar que, a cada milésimo de segundo, são inseridos um volume enorme de dados na rede mundial de computadores. Assim sendo, o mundo atual dos dados digitalizados contém uma inimaginável quantidade de informações disponíveis, base que está em crescimento constante e crescente (DAVENPORT; BARTH; BEAN, 2012).

Chen, Chiang e Storey (2012) e Davenport (2014) distinguem três períodos do uso de sistemas de informação para a análise de dados como apoio aos negócios. Os marcos estão atrelados ao surgimento de inovações tecnológicas que revolucionaram o processamento das informações dentro e fora das organizações (Figura I).

Figura I - Linha do tempo da evolução dos sistemas de informação nas empresas



Fonte: Adaptado de Silveira, Marcolin e Freitas (2015), Chen *et al.* (2012), Agrawal (2014) e Davenport (2014).

O uso de ferramentas de tecnologia da informação pelas empresas é descrito desde a década de 50, quando a inteligência artificial para uso corporativo já era tema de estudos (CHEN; CHIANG; STOREY, 2012). Inicialmente, a tecnologia desenvolvida para ajudar a tomada de decisões nos negócios se resumia a sistemas e processadores que de mecânicos tornaram-se eletrônicos (AGRAWAL, 2014).

Nos anos 90, aplicações como o ERP (Enterprise Resource Planning) ganharam força na estruturação dos negócios. Esta aplicação contribuiu para a organização das empresas, além de ter gerado uma base de dados para análise e tomada de decisão estratégica (AGRAWAL, 2014). A partir deste momento, foram aprimoradas as ferramentas de consultas aos bancos de dados, como o processamento analítico online (OLAP) e o Business Performance Management (BPM), quando se popularizou o termo *Business Intelligence* (BI), como um termo ligado à extração, segmentação, classificação e análise de dados para a detecção de anomalias e modelagens preditivas, com diversas aplicações de suporte a tomada de decisão nas empresas. Esta fase, que se estende até meados do ano 2000, foi denominada de *Business Intelligence and Analytics* 1.0 (BI&A 1.0) (CHEN; CHIANG; STOREY, 2012).

A partir dos anos 2000, Davenport, Barth e Bean (2012) e Chen, Chiang e Storey (2012) citam o advento da internet e do comércio eletrônico como fonte de uma grande mudança de paradigma. Neste momento, uma grande quantidade de dados das mais variadas formas e fontes passou a ser gerada, sendo compartilhada pela rede mundial de computadores. Padrões de navegação na rede, fluxo de cliques, compras realizadas, além de conteúdo gerado pelo próprio usuário em mídias sociais, blogs ou plataformas colaborativas são alguns dos exemplos de atividades conectadas pela internet que passaram a canalizar um fluxo constante e crescente de dados (SILVEIRA; MARCOLIN; FREITAS, 2015). Chen, Chiang e Storey (2012), Agrawal (2014) e Davenport, Barth, Bean (2012) denominam essa fase como a versão 2.0 do uso do *Business Intelligence and Analytics* (BI&A 2.0) e reforçam o advento da internet e do *e-commerce* como alavancas para a produção de dados que podem ser analisados e aplicados aos negócios.

Um novo salto na produção de dados se consolidou através da intensa massificação do uso de dispositivos móveis conectados à internet, aliado a uma gama crescente de aplicativos que potencializaram os efeitos do BI&A 2.0, justificando sua reformulação para BI&A 3.0. A partir deste período, as transformações advindas de smartphones e da maciça utilização da computação em nuvem, assim como da conexão de diversos equipamentos na internet (internet das coisas), e as redes sociais, a produção de dados tornou-se inimaginável e em fluxo contínuo (DAVENPORT, 2014; CHEN; MAO; LIU, 2014), quando o termo *Big Data* se fortaleceu.

Segundo Manyka *et al.* (2011), existem cinco maneiras para se criar valor transformacional através do *Big Data*: criar transparência; expor variabilidade que permita a experimentação; segmentar populações para personalizar ações; substituir / apoiar a tomada de decisão com algoritmos automatizados; e inovar em modelos de negócios, produtos e serviços. O investimento em ferramentas de inovação tem sido uma estratégia das organizações na busca de um melhor desempenho dos negócios e obtenção de vantagem competitiva (BRETHERTON *et al.*, 2011).



Um dos maiores desafios para o uso destas ferramentas está na análise destes dados e na compreensão de como eles podem trazer um diferencial competitivo nos negócios. Os dados captados no ambiente externo às empresas podem complementar as informações internas para melhor compreender clientes, mercados, concorrentes, produtos, ambiente de mercado, impacto de tecnologias ou até mesmo fornecedores (MARCHAND; PEPPARD, 2013; MAYER-SCHONBERGER; CUKIER, 2013).

Segundo pesquisa PWC (2017), executivos buscam tomar decisões em menor tempo, mas admitem que as organizações não estão prontas para isso. Já a consultoria Gartner divulgou que apenas 15% das organizações realmente implantaram seus projetos de big data em 2015 (CIO, 2016).

A análise e a aplicação do *Big Data* tornam-se um grande desafio, uma vez que, para prover informações que auxiliem as tomadas de decisões dos gestores, deve-se processar corretamente a grande quantidade de dados que devem ser consistentes e estarem disponíveis para gerarem inovação e conhecimento (ZHANG; CHEN; LI, 2013). Os dados estão sendo criados a taxas sem precedentes, o que requer que o *Big Data* tenha novas plataformas para a sua integração e para o seu processamento. A exploração e análise de dados através do *Big Data Analytics*, é feita com o uso de diferentes ferramentas, baseadas em análise preditiva, mineração de dados, estatística e inteligência artificial dentre outras (BRETERNITZ; LOPES, 2015).

A questão mais relevante talvez seja o de converter este grande volume de dados em algo que venha gerar informação relevante para as organizações, de maneira a gerar alguma vantagem competitiva (ERICKSON; ROTHBERG, 2014).

Encontrar pessoas com habilidades para analisar o *Big Data* é outro grande desafio. As competências exigidas são complexas e não exclusivamente orientadas para a tecnologia. Os profissionais necessitam habilidades múltiplas e entendimento dos processos na organização como um todo. As empresas recebem muitos dados e tem pouco tempo de processá-los e transformá-los em conhecimento, cabendo a elas a análise dos dados e dizer o que fazer com eles (GERMANN; LILIEN; RANGASWAMY, 2013).

É preciso também perceber que o *Big Data Analytics* possa, isoladamente, dar o real entendimento da informação e gerar conhecimento com objetividade e acurácia (COULDRY, 2014; BOYD; CRAWFORD, 2012). O uso da ferramenta é subjetivo, além de requerer saber coletar dados, agregá-los e fazer cruzamentos coerentes entre eles. Para tanto, as organizações devem se preparar, o que inclui pessoas, tecnologia e cultura (GERMANN; LILIEN; RANGASWAMY, 2013; BOYD; CRAWFORD, 2012; ROSS; BEATH; QUAAADGRAS, 2013). Outra limitação com relação ao *Big Data*, segundo Zhang et al. (2013), são os problemas relacionados à segurança durante análise de dados, de forma que eles possam ser comparados, se mantenham íntegros e não se percam com o tempo.

Segundo Taurion (2013), Manyika *et al.* (2011) e Diebold (2012), todos os setores de negócios serão afetados pelo *Big Data*, em maior ou menor grau, visto que extrair informação de grandes volumes de dados disponíveis na rede mundial de computadores será essencial para lidar com a concorrência e para o crescimento das empresas.

## ASPECTOS METODOLÓGICOS

O presente estudo caracteriza-se por ser de natureza quali-quantitativa, uma vez que usa simultaneamente abordagens qualitativa e quantitativa. É qualitativo por tratar-se de um estudo de caso único. Tem características quantitativas porque foram usadas estatísticas descritivas sobre os resultados coletados em fontes de dados documentais sobre as quais adotou-se análise de conteúdo (GODOY, 1995). Foi analisado o projeto “Perfil de risco”, realizado no ano de 2019, em uma grande operadora de saúde suplementar.

A operadora estudada foi selecionada pelo critério de acessibilidade à documentação e aos dados de pesquisa, assim como pela sua importância no contexto nacional no que tange à representatividade de sua carteira de clientes.

O custo com processos judiciais, liminares e processos administrativos junto ao órgão regulador (ANS) fazem parte do escopo do projeto e tem grande importância no equilíbrio econômico financeiro da empresa em questão, fazendo parte da sua curva A de gastos (UNIMED FORTALEZA, 2020).

A operadora de planos de saúde, objeto de estudo, é uma empresa de medicina de grupo que desenvolve suas atividades em Fortaleza e região metropolitana, estando entre as 20 maiores operadoras do país e sendo a oitava maior do sistema Unimed em número de clientes (ANS, 2020a; UNIMED BRASIL, 2020). Como operadora local, emprega 3.143 funcionários e detém 26,5% do mercado com 340.775 beneficiários. Cerca de 59,79% da sua carteira refere-se a planos coletivos de atendimento médico e 40,21% individuais. O faturamento em 2019 atingiu R\$ 2,4 bilhões, representando crescimento de 12,48% em relação a 2018 (UNIMED FORTALEZA, 2020).

Quanto aos objetivos, a pesquisa é descritiva, utilizada quando o objetivo do pesquisador consiste em descrever situações, acontecimentos e fatos, ou seja, dizer como se comporta um fenômeno (SAMPLERI; COLLADO; LUCIO, 2013). Na concepção de Gil (2008), a pesquisa descritiva tem como principal objetivo descrever características de determinada população ou fenômeno, ou o estabelecimento de relações entre as variáveis. Uma de suas características mais significativas está na utilização de técnicas padronizadas de coleta de dados. Andrade (2008) destaca que a pesquisa descritiva se preocupa em observar os fatos, registrá-los, analisá-los, classificá-los e interpretá-los, e o pesquisador não interfere nos mesmos.

A realização deste estudo de caso envolveu um levantamento documental, a partir de dados coletados nos setores de ouvidoria, jurídico interno e gerência de ciência de dados no período de janeiro a dezembro de 2019. Conforme Cunha, Yokomizo e Bonacim (2010), a análise de conteúdo é uma técnica que pode ser aplicada na análise de documentos sendo realizada com a intenção de sistematizar de forma objetiva os dados desestruturados encontrados, de maneira a quantificá-los. Os dados foram analisados entre 2 e 20 de janeiro de 2020.

O Projeto de “Perfil de risco” foi desenvolvido com o objetivo de sinalizar a propensão de cada beneficiário da Unimed Fortaleza em abrir notificação de intermediação preliminar (NIP) e/ou liminar judicial contra ela no período de seis meses futuros em relação ao mês de referência.

Utilizou-se como base para o desenvolvimento do modelo um período de seis meses (julho a dezembro do ano de 2018). Os dados coletados foram organizados em um banco de dados em Microsoft Excel® e organizados segundo o Quadro 3. Os itens utilizados na pesquisa foram referentes a transações junto à operadora no que tange aos serviços assistenciais, interações com o serviço de atendimento ao cliente (SAC), interações com a ouvidoria da operadora, NIPs junto à ANS e liminares judiciais.

### Quadro 3 - Dados coletados

Transações	Quantidade de transações nos últimos 30, 60, 90, 120, 150 e 180 dias Quantidade de negativas nos últimos 30, 60, 90, 120, 150 e 180 dias Proporção de negativas nos últimos 30, 60, 90, 120, 150 e 180 dias Tempo desde a última transação
SAC	Quantidade de ligações ao SAC nos últimos 30, 60, 90, 120, 150 e 180 dias Quantidade de ligações ao SAC do tipo informação nos últimos 30, 60, 90, 120, 150 e 180 dias Quantidade de ligações ao SAC do tipo solicitação nos últimos 30, 60, 90, 120, 150 e 180 dias Quantidade de ligações ao SAC do tipo reajuste nos últimos 30, 60, 90, 120, 150 e 180 dias
Ouvidoria	Tempo desde a última reclamação Quantidade de ligações para a ouvidoria nos últimos 30, 60, 90, 120, 150 e 180 dias Tempo de tratativa na ouvidoria
NIP	Quantidade de NIPs abertas nos últimos 30, 60, 90, 120, 150 e 180 dias Tempo desde a última NIP
Liminares Judiciais	Quantidade de Liminares abertas nos últimos 30, 60, 90, 120, 150 e 180 dias Tempo desde a última liminar

Fonte: Elaborado pelos autores (2020).

A ferramenta foi desenvolvida no período de julho a dezembro do ano de 2018, utilizando a técnica de *Machine Learning* (ML) para o desenvolvimento do modelo probabilístico. ML é o estudo e a construção de algoritmos computacionais que podem aprender com dados (HASTIE; TIBSHIRANI; FRIEDMAN, 2011). Tal técnica detecta automaticamente padrões nos dados e, em seguida, usa essa descoberta em um modelo preditivo, ou seja, funções que permitam o mapeamento de dados em um conjunto de classes que possam ser utilizados para classificar novos conjuntos de dados desconhecidos (MURPHY, 2012). Todo esse volume de dados pode conter informações importantes capazes de auxiliar na tomada de decisão, gerando vantagem competitiva.

O *out-of-time* (OFT) é um método para validação da qualidade do modelo, verificando se a base OFT compartilha do mesmo comportamento da base de desenvolvimento por meio de estatísticas, tais como o KS (Kolmogorov-Smirnov) com o intuito de confirmar que a base OFT segue a mesma distribuição da base de desenvolvimento. Utilizou-se o período de um mês com as mesmas informações da base utilizada para o desenvolvimento.

No período de produção, são escorados todos os beneficiários da Unimed Fortaleza ativos no mês de referência, utilizando análise de *cut-off* (Figura 2).

**Figura 2** - Tempo de desenvolvimento da ferramenta versus aplicação (produção)



Fonte: Elaborado pelos autores (2020).

Na produção são escorados todos os beneficiários da Unimed Fortaleza ativos no mês de referência, utilizando análise de *Cut-off* divide-se os escores em cinco faixas: baixíssimo risco (até 499), baixo risco (entre 500 e 549), médio risco (entre 550 e 699), alto risco (entre 700 e 829) e altíssimo risco (acima de 830).

O risco refere-se à propensão em abrir NIP e/ou liminar contra a Unimed Fortaleza. Portanto, o menor risco se encontra na faixa de baixíssimo risco e a maior propensão se encontra na faixa de altíssimo risco.

Para o ajuste do modelo foi utilizado o algoritmo denominado *Random Forest* (Floresta Aleatória). O *Random Forest* é um dos modelos mais populares entre os de “árvores de decisão”. As árvores de decisão são estruturas gráficas organizadas em nós, ramos e folhas. Cada nó representa um teste em um valor de atributo, os ramos representam, cada um, um resultado do teste, e as folhas, classes ou distribuições de classe. A conversão do modelo gráfico das árvores em regras de decisão dá ao modelo um dos seus mais relevantes aspectos: a conversibilidade em regras (se antecedente, então consequente), o que permite uma melhor interpretação do modelo de decisão (LAROSE, 2005).

Ao contrário de outros algoritmos de *Machine Learning* baseados em técnicas estatísticas, a “árvore de decisão” é um modelo não paramétrico, não tendo pressupostos subjacentes. Entretanto, devemos ter cuidado ao identificar os problemas onde uma “árvore de decisão” é apropriada, e onde não é. A facilidade de interpretação e compreensão do algoritmo tem encontrado sua utilização em muitas aplicações que vão desde a agricultura, onde é possível prever as chances de chuva, dadas às diversas variáveis ambientais, até o desenvolvimento de softwares, onde é possível estimar o esforço de desenvolvimento, fornecendo os detalhes sobre os módulos. Ao longo dos anos, as abordagens baseadas em “árvores” evoluíram para um escopo muito mais amplo em aplicabilidade, bem como sofisticação. Elas estão disponíveis tanto no caso de variáveis discretas como de resposta mais complexa (HASTIE; TIBSHIRANI; FRIEDMAN, 2011; RAMASUBRAMANIAN, 2016).

As principais estatísticas do modelo provêm de uma matriz de dados denominada matriz de contingência (ou confusão) apresentadas a seguir (Quadro 4).

**Quadro 4:** Matriz de contingência

Modelo \ Observado	Não abriu NIP/liminar (%)	Abriu NIP/Liminar (%)
Não abriu NIP/liminar (%)	92% ( $n_{00}$ )	6% ( $n_{01}$ )
Abriu NIP/Liminar (%)	0,8% ( $n_{10}$ )	1.2% ( $n_{11}$ )

Fonte: Elaborado pelos autores (2020).

Legenda:  $n_{00}$ : frequência de zeros que realmente foram zero (acerto);  $n_{01}$ : frequência de zeros que na verdade foram um (erro);  $n_{10}$ : frequência de uns que na verdade foram zero (erro);  $n_{11}$ : frequência de uns que realmente foram um (acerto).

As principais medidas consideradas no ajuste do modelo são acurácia, sensibilidade e especificidade. Acurácia é a taxa de acerto do modelo, sendo definida por:  $(n_{00} + n_{11}) / (n_{00} + n_{01} + n_{10} + n_{11})$ . Sensibilidade é a taxa de verdadeiros positivos no modelo, sendo operacionalizada por:  $n_{11} / (n_{11} + n_{01})$ . Especificidade, por sua vez, representa a taxa de verdadeiros negativos no modelo, sendo calculada da seguinte forma:  $E = I - S$ .

Aplicando os modelos, foram obtidos os seguintes resultados. A acurácia foi 0,93, sensibilidade foi 0,17 e a especificidade foi 0,83.

Foi utilizado o teste de KS (Kolmogorov – Smirnov) para verificar se as duas amostras (desenvolvimento e validação OFT) seguem a mesma distribuição, isto é, se a distância entre as distribuições for significativamente próxima de 0.

Primando pela didática e visando facilitar o entendimento dos dados coletados, esses foram agrupados em tabelas e gráficos que se mostram expostas na seção seguinte.

## RESULTADOS E DISCUSSÃO

O projeto foi dividido em cinco fases. Foram elas: definição do problema, extração e integração de dados, armazenamento, análise e tomada de decisões.

A primeira fase é uma das fases mais cruciais do processo de desenvolvimento e implantação de uma solução de *Analytics*. É nessa fase que ocorre a definição do conceito e da proposta da solução baseada na necessidade do cliente. Nesse momento, foi realizado o mapeamento dos processos envolvidos e adotados pela área a ser atendida.

Após uma definição clara do problema a ser enfrentado, o fluxo de levantamento de requisitos foi iniciado, a fim de buscar todas as necessidades para o desenvolvimento da solução a partir do cenário vislumbrado. Foi realizada a análise das condições funcionais e técnicas para atender a demanda do setor e a definição dos detalhes técnicos sobre o funcionamento da solução.

De acordo com a necessidade da solução e as fontes dos dados necessárias, deu-se início à segunda fase. O desenvolvimento da solução teria que utilizar dados estruturados ou desestruturados, tornando necessária a adesão de uma ou mais ferramentas para a o tratamento ou pré-processamento dos dados de origem. Segundo Demchenko *et al.* (2013), dados estruturados contém uma organização para serem recuperados. É como se fossem etiquetas, linhas e colunas que identificam diversos pontos sobre aquela informação e tornam o trabalho da tecnologia mais simplificado. Quando não se consegue identificar uma organização clara dos dados armazenados, este é um dado não estruturado. Após a extração, os dados foram integrados.

A terceira fase foi iniciada, com o armazenamento em bancos de dados tipo SQL e Data Lake.

Na quarta fase, descobre-se a quantidade de dados nulos na base, a distribuição dos dados, o balanço médio da população e a semi-estruturação dos dados. Nesse momento realizamos: a) análise de

dispersão que indica a existência, ou não, de relações entre variáveis de um processo e sua intensidade, representando duas ou mais variáveis em função da outra; b) análise de *outliers* que é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise e; c) análise temporal a partir de uma sequência de números coletados em intervalos regulares durante um período de tempo.

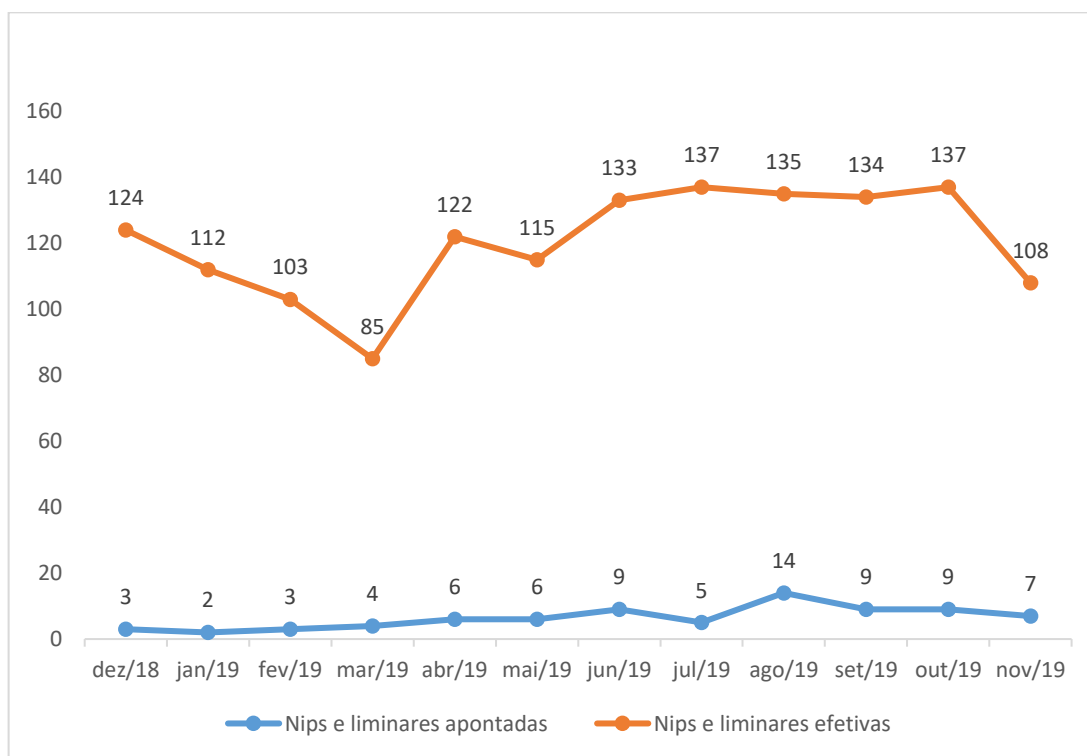
Nesse momento foram criados grupos de variáveis que são a matéria-prima das soluções de *Analytics*, pois são a partir delas que os modelos estatísticos são treinados, portanto, a acurácia de um modelo está intimamente dependente da capacidade de suas variáveis para justificarem um resultado. Um *book* de variáveis contém os dados em sua forma mais adequada para alimentar o modelo estatístico, após sofrerem todas as transformações e adequações necessárias. Alguns exemplos podem ser citados: Variáveis de frequência, temporais, proporcionais em relação ao tempo, quantitativas e qualitativas.

Em seguida, foi realizada a modelagem, o treinamento do modelo e a validação.

Na quinta e última fase, foi selecionada a ferramenta de *Business Intelligence* (BI) para acesso ao conhecimento gerado pelas ferramentas. Os modelos precisam ser monitorados e atualizados com frequência para que sejam analisadas possíveis mudanças nas variáveis. Com o tempo os modelos podem perder acurácia devido à mudança de comportamento, gerando defasagem entre a base de desenvolvimento e a base de conhecimento.

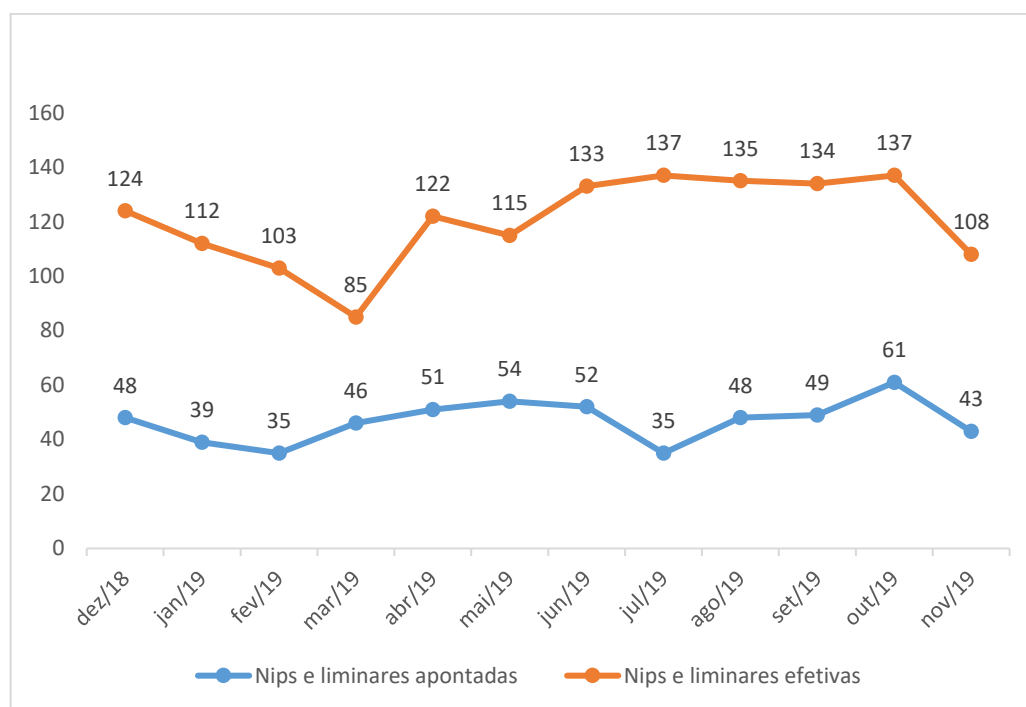
Concluídas as fases anteriores, foram analisados os casos de NIPs e liminares ocorridos após a entrega do modelo e a população foi cruzada com os scores de propensão do modelo estatístico. Quando são selecionados os beneficiários com perfil alto e altíssimo risco, a ferramenta sinaliza em média 5,2% do total de NIPs e liminares (Figura 3). No caso da seleção dos beneficiários com perfil médio, alto e altíssimo risco, há um incremento na sinalização para em média 39,5% (Figura 4).

**Figura 3** - NIPs e liminares apontadas e total de NIPs e liminares efetivas (alto e altíssimo risco).



Fonte: Elaborado pelos autores (2020).

**Figura 4** - NIPs e liminares apontadas e total de NIPs e liminares efetivas (médio, alto e altíssimo risco).



Fonte: Elaborado pelos autores (2020).

A tabela I evidencia a estratificação do perfil de risco no período de março a agosto de 2019. Nesse período, em média, 104 pacientes por mês foram sinalizados como alto e altíssimo risco, e 628 como médio, alto ou altíssimo risco.

**Tabela I** - Distribuição dos beneficiários por faixa de risco e período

Faixa de risco	mar/19	abr/19	mai/19	jun/19	jul/19	ago/19
1 - Altíssimo risco	0	2	1	2	11	41
2 - Alto risco	159	240	159	240	118	280
3 - Médio risco	1.516	1.851	2.442	1.851	829	1.568
4 - Baixo risco	936	1.185	1.547	1.185	959	1.262
5 - Baixíssimo risco	337.485	336.164	335.661	336.164	340.078	337.574
<b>Total geral</b>	<b>340.096</b>	<b>339.442</b>	<b>339.810</b>	<b>339.442</b>	<b>341.995</b>	<b>340.725</b>

Fonte: Elaborado pelos autores (2020)

A análise de grande volume de dados pode conter informações importantes capazes de auxiliar na tomada de decisão, gerando maior efetividade mercadológica (LAROSE, 2005). Para atingir tal efetividade é preciso ter a habilidade de extrair a informação ou o conhecimento útil dos dados coletados. É necessário também realizar de forma estruturada a confecção das fases de implantação e implementação das ferramentas, assim como da maneira como o conhecimento gerado será analisado e tratado. Fazer uso do *Big Data* e *Data analytics* vem se tornando uma necessidade de se manter competitivo no mercado (DAVENPORT, 2013), o que contribui para a sustentabilidade no longo prazo.

Os resultados apresentados evidenciam um nível de sinalização de potenciais geradores de NIPs e liminares judiciais em níveis bastante satisfatórios perfazendo um total de 39,5% quando são incluídos os níveis 1,2 e 3 do perfil de risco. Além da potencial redução em gastos com esses dois setores que no ano de 2019 geraram um custo de R\$ 69,8 milhões de reais (UNIMED FORTALEZA, 2020), ainda

há o ganho de imagem como consequência da resolução de problemas ainda em fases iniciais, evitando o desgaste dos beneficiários em face das operadoras de saúde suplementar.

Seja na área pública, reconhecidamente carente, ou na privada, em que as operadoras de planos de saúde reclamam de perdas no setor, não há dúvidas de que é preciso rever processos e investir em tecnologias capazes de aumentar o controle e melhorar a qualidade da assistência ao cliente (PINOCHET; LOPES; SILVA, 2014).

Segundo Aceto *et al.* (2018) o progresso em tecnologias da informação em saúde oferece novas oportunidades, bem como novos cenários de aplicação, trazendo benefícios em qualidade dos serviços médicos e na redução dos custos de saúde. Segundo os autores a literatura relacionada ao tema parece ser ainda muito dispersa e fragmentada, justificando novos estudos.

Os sistemas de tecnologia da informação se propõem a contribuir com a melhoria da qualidade do atendimento em saúde, aumentando a eficiência em diversos tipos de processos. Podemos afirmar que um sistema de TI deve servir para gerenciar a informação que os profissionais de saúde precisam para desempenhar suas atividades com efetividade, facilitar a comunicação, integrar a informação e coordenar as ações entre os diversos membros da equipe profissional de atendimento, e ao mesmo tempo fornecer apoio à gestão e à estratégia, inclusive sinalizando com processos inovativos ou até mesmo disruptivos.

## CONSIDERAÇÕES FINAIS

O cuidado com a saúde representa um importante desafio social e econômico que cada país enfrenta. O aumento do custo da assistência médica, envelhecimento e o constante crescimento populacional influenciam as soluções de saúde, aumentando as necessidades de novas e mais avançadas soluções científicas. Este cenário demanda a implantação de novas tecnologias em saúde, que contribuam com a segurança do paciente, auxiliem o processo de tomada de decisão ao médico, e que suportem o processo de gestão da saúde (GONÇALVES; MATOS; CHANG JUNIOR, 2019).

Segundo Parente e Van Horn (2006), o uso da TI na gestão da saúde aumenta a qualidade e a segurança do atendimento ao paciente, a eficiência geral do sistema e a redução de custos. Para Athey e Stern (2002) além da redução global de custos e da melhoria na qualidade dos serviços de saúde, a TI aumenta a precisão e reduz o tempo de atendimento diminuindo a probabilidade de mortalidade.

O presente trabalho buscou analisar a contribuição das ferramentas de *Big Data e Data Analytics* no contexto de uma operadora de saúde do sistema Unimed. A evolução destas ferramentas de inovação no mundo dos negócios propicia novas oportunidades na melhoria de processos e no entendimento de dados o que ajuda a direcionar as empresas ao alcance de seus objetivos estratégicos.

Esta análise sinaliza a possibilidade de atuação com medidas de relacionamento ao cliente ou modificação de fluxos, seja na esfera assistencial como na administrativa, a fim de minimizar os efeitos negativos desses importantes ofensores em custos, o que gera maior sustentabilidade no médio e longo prazo à operadora de saúde.

Pode-se afirmar com relação ao primeiro objetivo específico, qual seja, apresentar os passos para a implantação da ferramenta, que as ferramentas de *Big Data e Data Analytics* foram submetidas a um período de 6 meses de desenvolvimento entre julho e dezembro do ano de 2018, utilizando *Machine Learning* e seguindo as fases de: a) definição do problema; b) extração e integração de dados; c) Armazenamento; d) Análise e; e) tomada de decisão. A partir de então a base para consideração do perfil de risco era aplicada no mês seguinte.

Com relação ao segundo objetivo específico, qual seja, apresentar os resultados da ferramenta, percebe-se que a ferramenta consegue sinalizar uma quantidade significativa de clientes atingindo o nível de sinalização de 39,5% quando analisado o grupo de beneficiários de médio, alto e altíssimo risco.

Como sugestão para outros estudos no segmento da prestação de serviços de saúde, e com base nos aspectos observados durante o desenvolvimento desta pesquisa, sugere-se avaliar outras possíveis estratégias de utilização as ferramentas de *Big Data e Data Analytics* como a avaliação do perfil assistencial dos beneficiários e rede prestadora de serviços quando a ferramenta poderia sinalizar oportunidades de

otimizar a utilização da rede de assistência médica oferecida pela operadora, podendo trazer potencial redução em custos e melhoria da qualidade do atendimento prestado.

## REFERÊNCIAS

ACETO, G.; PERSICO, V.; PESCAPÉ, A. The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges. **Journal of Network and Computer Applications**, v. 107, 2018.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR - ANS. **Caderno de informação da saúde suplementar:** beneficiários, operadoras e planos. Dez. 2008. Disponível em: <<http://www.ans.gov.br/aans/noticias-ans/sobre-a-ans/906-caderno-de-informacao-da-saude-suplementar-edicao-dezembro2019>>. Acesso em: 3 abr. 2020a.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR - ANS. **Dados do setor.** Disponível em: <<http://www.ans.gov.br/aans/noticias-ans/numeros-do-setor/5348-saude-suplementar-fecha-2019-com-47-milhoes-de-beneficiarios-de-planos-de-saude>>. Acesso em: 3 abr. 2020b.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR - ANS. **Resolução normativa - RN nº 435**, de 23 de novembro de 2018 (Anexo I. c.I normas gerais p. II-12). Dispõe sobre o Plano de Contas Padrão da ANS para as operadoras de planos de assistência à saúde. Disponível em: <<http://www.ans.gov.br>>. Acesso em: 1 mai. 2019c.

AGRAWAL, D. Analytics based decision making. **Journal of Indian Business Research**, 6(4), 332–340, 2014.

ALBUQUERQUE, Ceres et al. A situação atual do mercado da saúde suplementar no Brasil e apontamentos para o futuro. **Ciência e Saúde Coletiva**, Rio de Janeiro, v. 13, n. 5, p. 1421-1430, out. 2008.

ANDRADE, M. M. **Como preparar trabalhos para cursos de pós-graduação:** noções práticas. 7. ed. São Paulo: Atlas, 2008.

ATHEY, S.; STERN, S. The impact of information technology on emergency health care outcomes. **Journal of Economics**, v. 33, n. 3, p. 399-432, 2002.

BAHIA, L. O mercado de planos e seguros de saúde no Brasil: tendências pós-regulamentação. In: NEGRI, B.; DIGIOVANNI, G. **Brasil - radiografia da saúde**. Campinas: Instituto de Economia da Unicamp, p.325-61, 2001a.

BAHIA, L. Planos privados de saúde: luzes e sombras no debate setorial dos anos 90. **Ciência e Saúde Coletiva** v.6 n.2, p. 329-339, 2001b.

BAHIA, L.; VIANA, A. Breve histórico do mercado de planos de saúde no Brasil. In: BRASIL. Ministério da Saúde. **Regulação e saúde:** estrutura, evolução e perspectivas da assistência médica suplementar. Rio de Janeiro: ANS, 2002.

BOYD, D.; CRAWFORD, K. Critical questions for big data. **Information, Communication & Society**, 15(5), 662-679, 2012.



BRASIL. **Lei nº 9.656**, de 03 de junho de 1998. Dispõe sobre os planos e seguros privados de assistência à saúde. Diário Oficial da União, Poder Executivo, Brasília, DF, 03 jun. 1998.

\_\_\_\_\_. **Lei nº 9.961**, de 28 de janeiro de 2000. Dispõe sobre a criação da Agência Nacional de Saúde Suplementar – ANS e dá outras providências. Diário Oficial da União, Poder Executivo, Brasília, DF, 29 jan. 2000.

\_\_\_\_\_. Ministério da Saúde. Secretaria-Executiva. Departamento de Economia da Saúde, Investimentos e Desenvolvimento. **Glossário temático: economia da saúde / Ministério da Saúde, Secretaria-Executiva, Departamento de Economia da Saúde, Investimentos e Desenvolvimento.** – 3. ed., 1. reimpr. – Brasília: Ministério da Saúde, 92 p., 2013.

BRETHERTON, P.; FERRER, M.; HYLAND, P.; SANTA, R.; SOOSA, C. The role of cross-functional teams on the alignment between technology innovation effectiveness and operational effectiveness. **International Journal of Technology Management**. 55(1-2), 2011.

BRETERNITZ, V. J.; LOPES, F. S.; SILVA, L. A. Big Data/Analytics: Formação e Gestão De Cientistas De Dados. In: 12th CONTECSI, 2015, São Paulo. **Anais do 12th CONTECSI**, 2015.

CHEN, H.; CHIANG, R. H.; STOREY, V. C. Business Intelligence and Analytics: From Big Data to Big Impact. **MIS quarterly**, 36(4), 1165– 1188, 2012.

CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile Networks and Applications**, 19(2), 171–209, 2014.

CHEN, C. P.; ZHANG, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. **Information Sciences**, 275, 314–347, 2014.

CHOW-WHITE, P. A.; GREEN, S. Data Mining Difference in the Age of Big Data: Communication and the social shaping of genome technologies from 1998 to 2007. **International Journal of Communication**, 7, 28, 2013.

CIO (2016). Recuperado de <http://cio.com.br/tecnologia/2016/12/01/menos-empresas-estao-investindo-mais-em-big-data-afirma-gartner>>. Acesso em: 3 abr. 2020.

COULDRY, N. A necessary disenchantment: Myth, agency and injustice in a digital world. **The Sociological Review**, 62(4), 880-897, 2014.

CUNHA, J. A. C.; YOKOMIZO, C. A.; BONACIM, C. A. G. Miopias de uma lente de aumento: as limitações da análise de documentos no estudo das organizações. In: ENCONTRO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 34. **Anais...** Rio de Janeiro: ANPAD, 2010.

DAVENPORT, T. H. How strategists use “big data” to support internal business decisions, discovery and production. **Strategy and Leadership**, 42(4), 45–50, 2014.

DAVENPORT, T. H.; BARTH, P.; BEAN, R. How “big data” is different. **MIT Sloan Management Review**, 54(1), 2012.

DAVENPORT, T. H. Analytics 3.0. **Harvard Business Review**, Dec., 64-72, 2013.

DEMCHENKO, Y.; GROSSO, P.; DE LAAT, C.; MEMBREY, P. Addressing Big Data issues in scientific data infrastructure. **Colaboration Technologies ans Systems (CTS)**, 2013.

DIEBOLD, Francis. **On the Origin (s) and Development of the Term “Big Data”**. Pennsylvania: Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012.

DUARTE, A. L. C. M.; OLIVEIRA, F. M.; SANTOS, A. A.; SANTOS, B. F. C. Evolução na utilização e nos gastos de uma operadora de saúde. **Ciência e Saúde Coletiva**. v.22, n.8, p. 2753-2762, 2017.

ERICKSON, S.; ROTHBERG H. Big Data an Knowledge: Establishing a conceptual foundation. **The Electronic Journal of Knowledge Management**. 12(2) 108-116, 2014.

GANDOMI, A. HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, p. 137–144, 2015.

GERMANN, F.; LILIEN, G. L.; RANGASWAMY, A. Performance implications of deploying marketing analytics. **International Journal of Research in Marketing**, 30(2), 114-128, 2013.

GIL, A. C. **Métodos e técnicas da pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de Administração de Empresas**, v. 2, n.35, p. 57-63,1995.

GONÇALVES, M.; MATTOS, C. A.; CHANG JUNIOR, J. Fatores Críticos de Adoção da Tecnologia da Informação (TI) em Saúde e o seu Impacto na Gestão: um estudo exploratório. **Revista Gestão da Produção Operações e Sistemas**, [S.l.], v. 14, n. 3, p. 209, jul. 2019.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. **The elements of statistical learning: data mining, inference, and prediction**. New York, Springer, 2011.

IESS – INSTITUTO DE ESTUDOS DE SAÚDE SUPLEMENTAR. **Variação de Custos Médico-Hospitalares: Sumário Executivo**. Edição: fev. 2020 Data-base: junho/2019. Disponível em: <<https://www.vcmh.com.br/>>. Acesso em: 11 abr. 2020.

LAROSE, D. T. **Discovering knowledge in data**. An introduction to data mining. New Jersey, EUA: John Wiley & Sons, Inc, 2005.

LEAL, R.M.; MATOS, J.B.B. Planos de saúde: uma análise dos custos assistenciais e seus componentes. **Revista de Administração de Empresas (RAE)**, [S.l.], v. 49, n. 4, p. 447-458, out. 2009.

MCAFFEE, A.; BRYNJOLFSSON, E. The Big Data: The Management Revolution. **Harvard Business Review**. p. 2-9. Oct, 2012.

MALTA D.C.; CECILIO L.C.O.; MERFHI E.E.; FRANCO T.B., JORGE A.O.; COSTA M.A. Perspectivas da regulação na saúde suplementar diante dos modelos assistenciais. **Ciência e Saúde Coletiva**, v. 7, n. 1, p. 61-66, 2004.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J., DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data: The next frontier for innovation, competition, and productivity**. 2011. Disponível em: [www.mckinsey.com/mgi](http://www.mckinsey.com/mgi). Acesso em: 30 dez 2019.

MARCHAND, D. A.; PEPPARD, J. Why IT fumbles analytics. **Harvard Business Review**, 91(1), 104–112, 2013.

MAYER-SCHONBERGER, V.; CUKIER, K. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana (Vol. 1). Elsevier Brasil, 2013.

MIRANDA, C.R. Gerenciamento de custo em plano de assistência à saúde. In: Agência de Saúde Suplementar. **Documento técnico de apoio ao Fórum de Saúde Suplementar de 2003** (Tomo 2). Rio de Janeiro: Bookmark, p. 425-461, 2004.

MURPHY, K.; **Machine Learning**: A Probabilistic Perspective. MIT Press, Cambridge, MA, 2012.

OCKÉ-REIS, C.O.; ANDREAZZI, M.F.S.; SILVEIRA, F.G. O mercado de planos de saúde no Brasil: uma criação do Estado? **Rev. Econ. Contemp.**, Rio de Janeiro, v. 10, n. 1, p. 157-185, 2006.

OLIVEIRA, S.B. **A judicialização da saúde suplementar e a necessidade de justiça especializada**, 2011. Disponível em < [http://www.jurisway.org.br/v2/dhall.asp?id\\_dh=6739](http://www.jurisway.org.br/v2/dhall.asp?id_dh=6739)>. Acesso em 26 dez 2019.

PARENTE, S. T.; VAN HORN, L. Valuing hospital investment in information technology: does governance make a difference? **Health Care Financing Review**, v. 28, n. 2, 2006.

PIETROBON, L.; PRADO, M. L.; CAETANO, J. C. Saúde suplementar no Brasil: o papel da Agência Nacional de Saúde Suplementar na regulação do setor. **Physis**, Rio de Janeiro, v. 18, n. 4, p. 767-783, 2008.

PINOCHET, L. H. C.; LOPES, A. S.; SILVA, J. S. Inovações e tendências aplicadas nas tecnologias de informação e comunicação na gestão da saúde. **Revista de Gestão em Sistemas de Saúde**, v. 3, n. 2, p. 11-29, 2014.

RAMASUBRAMANIAN, K., Singh, A.: Machine learning theory and practices. In: **Machine Learning Using R**, New Delhi, India, 24 Dec., 2016.

RONCALLI, A.G. O desenvolvimento das políticas públicas de saúde no Brasil e a construção do Sistema Único de Saúde. In: Antonio Carlos Pereira (Org.). **Odontologia em Saúde Coletiva**: planejando ações e promovendo saúde. Porto Alegre: Artmed, Cap. 2, p. 28-49, 2003.

ROSS, J. W.; BEATH, C. M.; QUAADGRAS, A. You may not need big data after all. **Harvard Business Review**, 90-98, Dec., 2013.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, P. B. **Metodologia de pesquisa**. 5. ed. São Paulo: McGraw-Hill, 2013.

SALAZAR, A.L.; RODRIGUES, K.; NUNES JÚNIOR, V.S. Assistência privada à saúde: regulamentação, posição do IDEC e reflexos no sistema público. In: BRASIL/MS. **Direito sanitário e saúde pública**, v. 1, p. 333-362, 2005.

SILVEIRA, M.; MARCOLIN, C. B.; FREITAS, H. M. R. Uso Corporativo do Big Data: Uma Revisão de Literatura. **Revista de Gestão e Projetos - GeP**, [S.l.], v. 6, n. 3, p. 44-59, 2015.

SOUSA, M.A; SALVALAIO, D. Rede própria ou rede credenciada: uma análise comparativa de custos em uma operadora brasileira de planos de saúde. **Revista Panam Salud Publica**, v.28, n.4, p.305-310, 2010.

TAURION, C. **Big Data**. Rio de Janeiro: Brasport, 2013.

TAVEIRA, M. Controle de custos em saúde: redução a qualquer preço ou racionalização na busca da eficácia? Elementos para discussão. **Saúde Debate**. 23:68-80, 1999.

UNIMED FORTALEZA. **Relatório de Gestão e sustentabilidade 2019**. Disponível em: <<https://www.unimedfortaleza.com.br/relatorios-de-gestao>>. Acesso em 10 de janeiro de 2020.

UNIMED BRASIL. **Panorama da Saúde Suplementar no Brasil**. Março de 2020. Disponível em <[https://www.unimed.coop.br/portalunimed/flipbook/unimed\\_brasil/panorama\\_saude\\_em\\_numeros\\_março\\_2020/](https://www.unimed.coop.br/portalunimed/flipbook/unimed_brasil/panorama_saude_em_numeros_março_2020/)>. Acesso em: 7 abr. 2020.

ZHANG, J.; CHEN, Y.; LI, T. Opportunities of innovation under challenges of big data. Proceedings - 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery, **FSKD**, 669–673, 2013.