

*MACHINE LEARNING*: APRENDIZAGEM DE MÁQUINA E  
MINERAÇÃO DE DADOS NO COMBATE A EVASÃO<sup>1</sup> FISCAL

**Resumo:** O presente artigo trata das técnicas de mineração de dados, aprendizagem de máquina e emprego de Inteligência Artificial na análise de dados visando a descoberta de padrões, detecção de comportamentos que apontem inconsistências, divergências e indícios de sonegação fiscal. Os sistemas transacionais atuais armazenam informações com capacidade superior à condição humana de manipulação e extração de conhecimento. Somados aos sistemas transacionais têm-se uma enxurrada de informações que são coletadas por meio de instrumentos (IOT – Internet das Coisas), redes sociais, e-mails e outros mecanismos geradores de dados que estão disponíveis para coleta que podem ser trabalhados para a obtenção de informações valiosas a fiscalização e gestão das receitas e despesas no âmbito da Administração Tributária. Nesse contexto, surge a necessidade de preparação de sistemas e recursos tecnológicos, bem como a capacitação técnica dos auditores e gestores para interpretar e analisar os dados, bem como, aplicar as técnicas avançadas de análise de dados visando a descoberta de padrões e extração de conhecimento útil tomada de decisão e atuação da fiscalização tributária de forma inteligente, abrangente, rápida e eficaz. A aplicação dessas técnicas avançadas de análises de dados e de auditoria em arquivos digitais se mostra a tendência do setor de fiscalização tributária e toda a administração tributária moderna. O artigo busca alertar para a necessidade urgente de mudança na forma de gestão para uma cultura orientada a dados buscando melhorar a eficiência da máquina pública em todos os setores e em especial a administração tributária que é setor responsável pelas receitas que irão proporcionar ganhos de eficiência e qualidade na prestação dos serviços públicos.

**Palavras Chave:** Internet das Coisas, Mineração de Dados, Aprendizagem de Máquina, Inteligência Fiscal. Escrituração Contábil Digital. Escrituração Fiscal Digital. Inteligência Artificial.

## 1 INTRODUÇÃO

---

<sup>1</sup> Evasão é o ato omissivo ou comissivo, de natureza ilícita, praticado com o fim único de diminuir ou eliminar a carga tributária, ocultando o verdadeiro ato ou a real situação jurídica do contribuinte. (FERRAGUT, 2014, p. 5).

São modalidades de evasão o *dolo* (sonegação omissiva), a *fraude material* (falsificação de registros, livros, documentos, dados, operações, balanços, despesas, receitas etc.) e a *simulação absoluta* ou *relativa* mediante a utilização abusiva das formas jurídicas de Direito Privado (abuso de formas= dissimulação relativa = fraude à lei fiscal). (COÊLHO, 2003, p. 193).

Com advento das novas tecnologias e aprimoramento das técnicas de escrituração, adentramos na era digital, onde as obrigações tributárias do contribuinte são entregues ao fisco em formato eletrônico, se por um lado trouxe simplificação da escrituração e entrega das obrigações tributárias exigidas das empresas e pessoas físicas, por outro, trouxe ao fisco a complexidade e a necessidade de tratar esse grande volume de dados gerados a partir dessas informações socioeconômicas e fiscais dos contribuintes e de igual forma extrair essas informações que são geradas nesses sistemas transacionais e aquelas provenientes de sistemas externos, mídias sociais, internet das coisas, e-mails e outras fontes de dados.

Este é o grande desafio da Administração Pública e aqui em especial tratamos da Administração Tributária, que para se adequar a essa nova realidade terá que investir pesadamente e continuamente no emprego de novas tecnologias de análises e tratamento de dados adequadas para carregar, armazenar, limpar e extrair e gerar conhecimento a partir dessas informações, fato este que já vem sendo percebido por grandes empresas e alguns órgãos do setor público que tem investido em aplicações do ecossistema *Hadoop/Big Data*, em Ferramentas de Análises de Dados (*Analytics* e *Business Intelligence*), e Ferramentas de Mineração de Dados na busca de eficiência e obtenção de vantagem competitiva.

Toda essa imensidão de conteúdo que é gerada nos mais diversos formatos (dados estruturados em colunas ou não estruturados) requer disponibilidade tecnológica para recepção, armazenamento, limpeza e tratamento.

Para tanto há no mercado softwares específicos que se encarregam da extração, carregamento, e limpeza de dados e tecnologias de armazenamento de grandes volumes de dados (*Big Data/Hadoop* e *Data Warehouse*) e de softwares específicos de mineração de dados (*Weka*, *Oracle Data Mining*, Linguagem R, *Python*, entre outras) ou do emprego de técnicas avançadas de aprendizagem de máquina (Algoritmos de Classificação, Associação e Agrupamento, Árvore de Decisão e Redes Neurais) e inteligência artificial (Agentes Inteligentes e Processamento de Linguagem Natural), sendo o emprego dessas ferramentas e técnicas de mineração de dados o foco deste artigo.

A aplicação de tais ferramentas compõe o aparato tecnológico de inteligência fiscal<sup>2</sup> necessário ao aprimoramento das técnicas de auditoria em arquivos eletrônicos, cruzamentos de informações socioeconômicas e tributárias a partir de base de dados, identificando inconsistências, indícios de irregularidades e identificando por mineração de dados padrões e correlações que apontam para indícios de fraudes e sonegação fiscal e no âmbito de gestão tais aplicações visam prever por meio de análises preditivas o comportamento das receitas e obter informações relevantes a tomada de decisão.

## **2 AS DIFERENTES NOMENCLATURAS: INTELIGENCIA ARTIFICIAL, INTELIGENCIA COMPUTACIONAL, APRENDIZAGEM DE MÁQUINA E COMPUTAÇÃO NATURAL.**

A literatura está repleta de nomenclaturas para as diversas técnicas de soluções de problemas e algoritmos computacionais e esses métodos são aplicados por pesquisadores, empresas, consultores e outros profissionais que atuam na área de inteligência de negócios e de mineração de dados. (Silva, et al., 2016).

A diversidade de técnicas e gama de profissionais que se valem dessas soluções faz com que surjam nomenclaturas distintas para contextos muitas vezes comuns, gerando uma certa confusão e dificuldade de entendimento. (Silva, et al., 2016).

Neste contexto cabe traçar distinção conceitual dessas terminologias, a fim de identificar e compreender os métodos, ferramentas e as suas aplicações.

---

<sup>2</sup> [...] entende-se por Inteligência fiscal a atividade de reunir informações com a finalidade de assessorar a administração tributária no planejamento e execução de suas atribuições legais, particularmente no combate à sonegação fiscal, à evasão fiscal, ao contrabando e ao descaminho. A atividade de Inteligência fiscal assume características proativas e reativas. Quando proativa, estará relacionada com a atividade de obtenção e análise de informações, possibilitando a elaboração de prognósticos na área tributária, visando ao combate a ilícitos e infrações tributárias ou aduaneiras, bem como ao aperfeiçoamento da legislação tributária ou de procedimentos administrativos. Quando reativa, poderá assumir a forma de investigação, auxiliando de forma decisiva na obtenção de provas e de indícios que poderão, oportunamente, ser utilizados tanto para fins administrativos quanto para instruir a representação criminal. (CIAT, 2006, p.6) *apud* (Leony, 2006, p. 67).

### 3 DEFINIÇÃO DE *MACHINE LEARNING* – APRENDIZAGEM DE MÁQUINA.

Dentre as mais diversas definições o aprendizado de máquina pode ser compreendido como a criação e uso de modelos que são aprendidos a partir de dados. (Grus, 2016).

A aprendizagem de máquina se vale de dois paradigmas de aprendizagem a saber:

*Aprendizado supervisionado e aprendizagem não supervisionada. Aprendizagem supervisionada baseia-se em um conjunto de objetos para os quais as saídas são conhecidas, ou em algum outro tipo de informação que represente o comportamento que deve ser apresentado pelo sistema. (Silva, et al., 2016, p. 16).*

*Aprendizado não supervisionado baseado apenas nos objetos da base, cujos rótulos são conhecidos. Basicamente, o algoritmo deve aprender a 'categorizar' ou rotular os objetos. (Silva, et al., 2016, p. 16).*

Para T. Mitchell a aprendizagem de máquina (AM) é área de pesquisa que visa desenvolver programas computacionais capazes de automaticamente melhorar seu desempenho por meio da experiência. *apud* (Silva, et al., 2016).

A aprendizagem de máquina tem como foco extrair informação a partir de dados de maneira automática e está ligada a mineração de dados, a estatística, à inteligência artificial e à teoria da computação, além de outras áreas como computação natural, sistemas complexos adaptativos e computação flexível, sendo seus principais métodos aqueles que trabalham com dados nominais, como árvores de decisão, as regras de associação e classificação, tabelas de decisão e outros. (Silva, et al., 2016).

A principal característica dessas técnicas de aprendizagem de máquinas está relacionada a capacidade de aprender adaptando-se ao ambiente de acordo com as regras preexistentes, alterando seu desempenho ao longo do tempo e no contexto da mineração de dados, a aprendizagem ou treinamento corresponde ao processo de ajuste e construção do modelo usando um mecanismo de apresentação ou uso dos objetos de base de dados. (Silva, et al., 2016).

#### 4. DEFINIÇÃO DE IA - INTELIGÊNCIA ARTIFICIAL

O termo IA - Inteligência Artificial foi cunhado pela primeira vez logo após a Segunda Guerra Mundial, mais precisamente em 1956, o campo da inteligência artificial tenta não apenas compreender, mas também construir entidades inteligentes e abrange uma variedade de aplicações que vai do geral a aplicações bem específicas tais como: jogos de xadrez, demonstração de teoremas matemáticos, criação de poesia, direção de um carro em estrada movimentada e diagnósticos de doenças. (Russel & Norvig, 2013).

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

Fonte: adaptado de Russel & Norvig, 2013, Inteligência Artificial, p. 4

Os sistemas de IA se tornaram tão comuns em aplicações na Web que o sufixo "bot" passou a integrar a linguagem cotidiana, nesse contexto destaca-se as aplicações das tecnologias de IA como base a muitas ferramentas da internet, como mecanismos de pesquisas, sistemas de recomendação, agregadores de conteúdo de construção de sites, assistentes virtuais, entre outras aplicações. (Russel & Norvig, 2013).

Até pouco tempo atrás não havia dados suficientes em volumes consideráveis que permitissem a aplicação com grande margem de precisão dos modelos de aprendizagem de máquinas e de implementação de IA.

A imensidão de dados e as novas tecnologias de armazenamento hoje possibilita tais aplicações e estudos mostram que técnicas como estas tem um desempenho ainda melhor quando aplicado a uma base maior de dados e permite que os modelos aplicados de algoritmos tenham uma maior acurácia quando aplicados em quantidade maior de dados. (Russel & Norvig, 2013).

## 5. DEFINIÇÃO DATA MINING – MINERAÇÃO DE DADOS

Segundo (Barbieri, 2011) os conceitos de mineração de dados ou garimpagem estão relacionados a tendência de buscar as correlações escondidas em altos volumes de dados que nem sempre estão acessíveis, em especial, quando se tratar de extrair informações de sistemas cotidianos.

“Minerar dados consiste em utilizar técnicas automáticas de estatística e de inteligência artificial para explorar grandes massas de dados à procura de relações ou padrões úteis para as organizações” (CARVALHO, 2005) *apud* (Andrade, 2009, p. 22).

*Data mining* é usado para reduzir comportamento fraudulento, particularmente em reivindicações de seguro e uso de cartão de crédito (CHAN et al., 1999) *apud* (Turban, et al., 2009) para identificar padrões de compra do cliente (HOFFMAN, 1999) *apud* (Turban, et al., 2009).

Tabela 2 - Funções de *data mining*, algoritmos e exemplos de aplicação

Função de <i>data mining</i>	Algoritmo	Exemplos de Aplicação
Associação	Estatística, teoria dos conjuntos	Análise de cesta de supermercado
Classificação	Árvore de decisão, redes neurais, controle, avaliação de risco, regras	<i>Marketing</i> direcionado, qualidade
Descoberta de sequencia	Estatística, teoria dos conjuntos	Análise de cesta de mercado ao longo do tempo, análise do ciclo de vida do cliente
Modelagem	Regressão linear e não linear, ajuste da curva, redes neurais	Previsão de vendas, nível de interesse, predição, controle de estoque
<i>Drill-down</i> e visão agregada dos dados	Visualização, usando muitas abordagens diferentes	Quase todas as aplicações anteriores

Fonte: adaptado de J.P Bigus, *Data Mining with Neural Networks*, McGraw-Hill, New York, 1996 *apud* Turban, et.al., 2009, p.157

O processo de mineração de dados pode ser dividido em tarefas de análises preditivas e descritivas, as tarefas preditivas usam os valores dos atributos descritivos para prever valores futuros ou desconhecidos de outros atributos de interesse, já as tarefas descritivas têm o objetivo de encontrar padrões que descrevem de maneira que o ser humano possa interpretar. (Silva, et al., 2016).

## 6. PRINCIPAIS TAREFAS DA MINERAÇÃO DE DADOS: ASSOCIAÇÃO, CLASSIFICAÇÃO, AGRUPAMENTO (CLUSTER), ESTIMAÇÃO, PREDIÇÃO, SUMARIZAÇÃO E DETECÇÃO DE ANOMALIAS (OUTLIERS).

De acordo com (Castro & Ferrari, 2016) a predição é uma terminologia usada para referir à construção e ao uso de um modelo para avaliar a classe de um objeto não rotulado ou para estimar o valor de um ou mais atributos de dado objeto, no primeiro caso tem-se a classificação e no segundo a regressão ou estimação. Exemplos das tarefas de Classificação incluem a identificação de *spams*, classificação de objetos, atribuição de crédito e detecção de fraudes.

Para (Silva, et al., 2016) a tarefa de predição consiste na análise de um conjunto de dados nos quais estão presentes os dados, descritos por atributos e seus rótulos associados, tem por objetivo descobrir um modelo capaz de mapear corretamente cada um dos dados aos seus rótulos, objetivo que é alcançado por meio da técnica de aprendizagem supervisionada.

A tarefa de agrupamento consiste no processo de separar um conjunto de objetos em grupos (*cluster*) de objetos similares e é utilizado para identificar tais grupos e, portanto, cada grupo formado pode ser visto como uma classe de objetos e considera dados de entradas não rotulados, ou seja, os dados de entradas não são conhecidos *a priori* esse processo é conhecido como treinamento não supervisionado ou aprendizagem não supervisionada. (Castro & Ferrari, 2016).

A análise por associação corresponde à descoberta de regras de associação que apresentam valores de atributos que ocorrem concomitantemente em uma base de dados sendo normalmente usadas em ações de *marketing* e para estudo de bases transacionais. (Castro & Ferrari, 2016).

Já as tarefas de Detecção de Anomalias (*Outliers*) tem aplicação em análise de base de dados que não seguem um padrão de comportamento ou não possuem a característica comum dos dados ou de um modelo que os represente. A maioria das ferramentas de mineração descarta as anomalias, entretanto, em algumas aplicações como na detecção de fraudes, os eventos raros podem ser mais informativos do que ocorrem regularmente. (Silva, et al., 2016).

Um bom algoritmo de detecção implica no descobrimento verdadeiro de anomalias evitando afirmações errôneas sobre desvios encontrados. Incluem a pesquisa por desvios temporais (mudanças significativas nos dados de séries

temporais) e desvios em grupos (diferenças não esperadas entre dois subconjuntos de dados) (Sumathi & Sivandam, 2006 *apud* (Levergger, 2013).

A tarefa de sumarização visa aplicar métodos para prover uma descrição compacta de um subconjunto de dados e é frequentemente utilizada no pré-processamento de dados para uma exploração intuitiva, quando valores inválidos são determinados através de métodos estatísticos como exemplo, a tabulação da média e desvio padrão, ou em casos mais sofisticados, através da distribuição de frequência de valores (Fayyad, Piatetsk-Shapiro e Smuty, 1996 *apud* (Levergger, 2013).

## **7. SISTEMA PÚBLICO DE ESCRITURAÇÃO DIGITAL – SPED: CT-e, ECD, ECF, EFD Contribuições, EFD ICMS IPI, EFD Reinf, e-Financeira, eSocial, MDF-e, NFC-e, NF-e e NFS-e.**

De acordo com o Manual de Autenticação de Livros digitais os objetivos do SPED são: a) promover a integração dos fiscos, mediante a padronização e compartilhamento das informações contábeis e fiscais, respeitadas às limitações legais; b) racionalizar e uniformizar as obrigações acessórias para os contribuintes com o estabelecimento de transmissão única de distintas obrigações acessórias de diferentes órgãos fiscalizadores; c) tornar mais célere a identificação de ilícitos tributários, com a melhoria do controle dos processos, a rapidez no acesso as informações e a fiscalização mais efetiva das operações com o cruzamento de dados e auditoria eletrônica. (Cleto & Oliveira, 2010, p. 12).

O Decreto nº 6.022, de 22 de Janeiro de 2007 em seu art. 2º, assim o define:

*O Sped é instrumento que unifica as atividades de recepção, validação, armazenamento e autenticação de livros e documentos que integram a escrituração contábil e fiscal dos empresários e das pessoas jurídicas, inclusive imunes ou isentas, mediante fluxo único, computadorizado, de informações.*

Os benefícios com a implantação do SPED de acordo com o Manual de Autenticação de Livros digitais são: a) redução de custos com a dispensa de emissão e armazenamento de documentos em papel; b) eliminação do papel; c)



redução de custos com a racionalização e simplificação das obrigações acessórias; d) uniformização das informações que os contribuintes presta as diversas unidades federadas; e) redução do envolvimento involuntário em práticas fraudulentas; f) redução do tempo despendido com a presença de auditores nas instalações do contribuinte; g) simplificação e agilização dos procedimentos sujeitos ao controle da administração tributária; h) fortalecimento do controle e da fiscalização por meio de intercâmbio de informações entre as administrações tributárias; i) rapidez no acesso às informações; j) aumento da produtividade do auditor através da eliminação dos passos para a coleta dos arquivos; l) possibilidade de troca de informações entre os próprios contribuintes a partir de leiaute padrão; entre outras. (Cleto & Oliveira, 2010, p. 12).

## **8 AUDITORIA EM ARQUIVOS ELETRONICOS COM O SOFTWARE CONTÁGIL**

A necessidade de desenvolvimento de uma ferramenta que possibilitasse tratar de forma eficiente um grande volume de informações contábeis fez com surgisse em meados de 2006 o *software* ContÁgil desenvolvido por um auditor fiscal da receita federal, inicialmente implantado em algumas unidades piloto, atualmente é adotada pela mesma como software padrão nas ações de auditoria em arquivos eletrônicos (Figueiredo, 2008).

O *software* foi desenvolvido com a finalidade de manipular arquivos de escrituração contábil digital objetivando automatizar as análises de auditoria fiscal em documentos fiscais disponibilizados pelo contribuinte nas ações fiscalização tributária e aduaneira por parte do auditor. (Figueiredo, 2008).

Dentre as facilidades da ferramenta destaca-se a) formatação de arquivos de extratos bancários para realização de procedimento de movimentação financeira incompatível; b) realização de “verificações obrigatórias” que comparam as informações da contabilidade com informações existentes em declarações oficiais; c) localização de contrapartidas de lançamentos contábeis. (Figueiredo, 2008).

Inicialmente a ferramenta foi concebida com funcionalidades que atendiam as análises de uma pequena unidade da receita federal, a idéia deu

muito certo que novas funcionalidades foram implementadas na ferramenta introduzindo complexas análises estatísticas e matemáticas. (Figueiredo, 2008).

O *software* ContÁgil realiza varredura nos lançamentos contábeis analisando através do método de partidas dobradas cada conta e sua contrapartida e apontando o resultado por meio de gráfico de fluxos contábeis com setas indicadoras das análises efetuadas. Partindo do Mapa apresentado caberá ao auditor fiscal mediante sua experiência e seu olhar apurado identificar as situações atípicas e suspeitas demonstradas no gráfico e solicitar documentação complementar para embasar as autuações. (Figueiredo, 2008).

Essa ferramenta foi objeto de assinatura de protocolo de cooperação nº 10/2015 assinado no X Encontro do ENAT firmado entre a Receita Federal, os Estado da Federação e a ABRASF para que os municípios possam utilizá-la em suas fiscalizações.

## **9 CRUZAMENTO DE INFORMAÇÕES EM ARQUIVOS DIGITAIS - SPED CONTÁBIL, SPED FISCAL, DCTF, NOTA ELETRONICA DE SERVIÇO, DECLARAÇÃO SERVIÇOS TOMADOS E OUTRAS OBRIGAÇÕES.**

Observa-se que a quantidade de obrigações tributárias a qual o contribuinte está obrigado a entregar ao fisco de todas as esferas é vasta e constitui um grande volume de informações socioeconômicas e fiscais de contribuintes estabelecidos no município que necessitam ser trabalhadas de forma ágil e eficaz, por meio de sistemas próprios de cruzamentos de informações, de detecção de fraudes e sonegação fiscal.

É evidente que para manipulação de arquivos digitais como SPED e DCTF que são obrigações tributárias entregues a Receita Federal há necessidade de firmar convenio ou no interesse da fiscalização em procedimento fiscal.

A formalização do convenio se justifica pela importância de se trabalhar com os arquivos diretamente na base de dados da Receita Federal, o compartilhamento de informações de interesse da fiscalização é autorizado pela Constituição federal conforme dispõe o art. 37, XXII:

*As administrações tributárias da União, dos Estados, do Distrito Federal e dos Municípios, atividades essenciais ao funcionamento do*

*Estado, exercidas por servidores de carreiras específicas, terão recursos prioritários para a realização de suas atividades e atuação de forma integrada, inclusive com o compartilhamento de cadastros e de informações fiscais, na forma da lei ou convênio.*

A troca de informações entre os entes de fiscalização constitui importante instrumento no combate à evasão fiscal e nesse ponto não cabe a alegação de quebra sigilo, pois não há quebra, mas somente a transferência de arquivos entre órgãos de fiscalização no interesse da arrecadação tributária.

Todas essas obrigações são passíveis de cruzamento no interesse do fisco em seus campos próprios, porém, para cruzar tais informações tributárias os fiscos têm investido na implementação de tecnologias capazes de extrair, carregar, limpar e armazenar com segurança os dados e informações econômicos-fiscais dos contribuintes e ferramentas adequadas a mineração de dados necessárias a descoberta padrões e correlações nos dados que possibilita identificar irregularidades, indícios de fraudes e sonegação.

## **10 CRUZAMENTO DE INFORMAÇÕES RELACIONADAS A OBRAS DE CONSTRUÇÃO CIVIL – SISOBRA-PREF**

A Secretaria da Receita Previdenciária - SRP do Ministério da Previdência Social em parceria com a DATAPREV desenvolveu o módulo para prefeitura cadastrar as obras, o sistema permite a padronização de relatório de alvarás e documentos de habite-se que os municípios estão obrigados a enviar mensalmente para a SRP em cumprimento ao artigo 50 da Lei 8212/91: "Para fins de fiscalização do INSS, o Município, por intermédio do órgão competente, fornecerá relação de alvarás para construção civil e documentos de "habite-se" concedidos".

A adesão ao referido sistema não é obrigatória, já que as prefeituras podem dispor de sistemas próprios de controle, neste caso conforme instrução normativa nº 1569, de 05 de junho de 2015 os municípios devem atender a entrega de acordo com leiaute estabelecido nesta Instrução Normativa e para transmissão de arquivos é necessária a adesão ao Domicílio Tributário Eletrônico – DTE.

A utilização do sistema SISOBRA- PREF constitui uma importante fonte de dados para criação de uma “malha fina” referente às obras de construção civil, no que tange aos recolhimentos de ISS e fonte de cruzamento com o sistema de gestão tributária e nota eletrônica a fim de monitorar os recolhimentos e orientar a fiscalização e combater a evasão fiscal que se verifica nesse importante setor.

## **11 CRUZAMENTO DE INFORMAÇÕES “MALHA FISCAL” EMPRESAS OPTANTES SIMPLES NACIONAL**

As recentes alterações trazidas pela Lei Complementar 155/2016 de que altera o limite para R\$4,8 milhões para empresas de pequeno porte a vigorar a partir de 1º de janeiro de 2018, juntamente com a inclusão de outras atividades denota a importância do regime simples nacional e o peso que este tem para os municípios, uma vez, cada vez mais empresas estarão amparadas pelo regime, exigindo-se por parte do fisco municipal e de outras esferas, um maior cuidado e atenção com a arrecadação tributária desse regime.

Sabe-se que o corpo de auditoria fiscal é na maioria das vezes insuficiente as demandas de fiscalização, quer seja para empresas de grande porte, ainda mais o será se considerarmos as empresas do regime simples nacional, como o número de empresas no regime aumenta consideravelmente de igual forma deve aumentar a preocupação do fisco em manter o controle das receitas e a fiscalização a fim de evitar perdas de receitas e perda de presença fiscal.

A inviabilidade de fiscalizar todas as empresas denota a importância de investir em aparato tecnológico de suporte a fiscalização, em se tratando de empresas do regime simples nacional, isso deve ser intensificado já que, os recursos devem ser direcionados, de forma planejada e pautada por critérios que denotem a importância de atuação da fiscalização, isso somente é possível mediante a criação de uma “malha fiscal” para essas empresas.

Nesse sentido mais uma vez destaca-se a importância de garantir investimentos em infraestrutura e ferramentas tecnológicas que permitam ao fisco os monitoramentos das obrigações tributárias, a fim de verificar com maior eficiência e agilidade o comportamento das receitas proveniente deste regime.

Destaca-se ainda a importância das ferramentas de *Business Intelligence* e outras aplicações que permitam ao fisco efetuar o cruzamento de informações proveniente das declarações do regime simples nacional com a nota eletrônica, possibilitando identificar possíveis irregularidades nas declarações e combater a evasão fiscal, também em procedimentos de fiscalização da escrituração contábil aplicar técnicas de mineração de dados a fim de detectar padrões que identifique ou apontem que há divergências, irregularidades ou indícios de sonegação fiscal.

## **12 TEXT MINING - MINERAÇÃO DE TEXTO EM ARQUIVOS XML DA NOTA ELETRONICA**

No âmbito da mineração de dados merece destaque as técnicas de mineração de textos que podem ser aplicadas na fiscalização e nos procedimentos de auditoria a fim de detectar padrões de correlações que possam apontar irregularidades, fraudes e indícios de sonegação fiscal.

*Text mining* é definido como “a aplicação de data mining em arquivos de textos não estruturados ou menos estruturados”. (Turban, et al., 2009, p. 177)

Em estudo de caso realizado por (Madeira, 2015) que enfatiza a aplicação das ferramentas de mineração de texto voltada à extração de informações do corpo da nota fiscal carioca, em especial o campo de “discriminação de serviços” informações relevantes à orientação da fiscalização e o combate à evasão fiscal no que tange as receitas de prestação de serviços de engenharia consultiva - acompanhamento e fiscalização da execução de obras de engenharia, arquitetura e urbanismo, em que tem sido declarados pelo contribuinte com objetivo de deslocar a tributação para fora do município do estabelecimento prestador.

Para (Madeira, 2015) a utilização do método tem como principal objetivo criar uma “malha fina” de empresas suspeitas de emitirem notas eletrônicas de forma incorreta e permitir com isso o deslocamento da tributação para outro município indevidamente, com a aplicação da ferramenta de mineração de texto que se vale de algoritmos de classificação e agrupamento é possível determinar de acordo com as informações declaradas pelo contribuinte no campo “Discriminação do Serviço” qual a natureza do serviço prestado, e identificar se

o serviço enquadra-se na regra de incidência que desloca a tributação para fora do município ou se trata de serviços que são devidos no estabelecimento prestador, com isso há possibilidade de atuação pontual da fiscalização no combate à sonegação.

### **13 USANDO A ANÁLISE PREDITIVA PARA PLANEJAMENTO DE AÇÕES E MONITORAMENTO DO COMPORTAMENTO DAS RECEITAS.**

Sabe-se que as Unidades de Fiscalização Tributária carecem de recursos humanos em quantidade suficiente a realização das auditorias e cada vez mais tem-se quantidade menor, de pessoas para realizar estas importantes tarefas, também a de se destacar que a modernização da escrituração fiscal e contábil das empresas para formato digital, simplificou as obrigações tributárias das empresas no que tangem ao formato de entrega, mas a complexidade permanece, face a quantidade de obrigações tributárias que estas tem perante os fiscos em todas esferas.

Considerando o cenário atual das grandes empresas, médias e até mesmo as pequenas torna-se cada vez mais dispendioso fiscalizar suas obrigações tributárias sem o auxílio de ferramentas que possibilitem manipular os dados de forma dinâmica e efetuar cruzamentos necessários as atividades de fiscalização.

Também no que tange a programação da fiscalização sem auxílio de ferramentas torna-se difícil determinar com precisão quais são as empresas que não estão cumprindo suas obrigações de forma adequada e que necessitam ser fiscalizadas.

É nesse ponto que entra a análise preditiva, por meio desta é possível cruzar os milhões de registros para identificar orientações promissoras a fim de tomar as melhores decisões relativas a auditoria, aumentando o sucesso da operação e recuperação do imposto devido, com a análise preditiva é possível determinar o resultado provável futuro para um evento ou probabilidade de uma situação ocorrer e identificar relações e padrões. (Turban, et al., 2009).

Outra variável importante para a fiscalização é a possibilidade de determinar com precisão as empresas que mudaram o domicílio tributário por questões tributárias, ou até mesmo por motivos relacionados a esta questão.

Com emprego da análise preditiva é possível analisar e determinar com precisão as razões das mudanças, prever possíveis perdas de receitas e de empresas, possibilitando ao fisco aprimorar as políticas de incentivos fiscais e de mecanismos de retenção, para que estas empresas obtenham vantagens competitivas e permaneçam instaladas no município.

A análise preditiva é usada para examinar grandes quantidades de dados com diferentes variáveis incluindo *clustering*, árvore de decisão, modelagem de regressão, redes neurais, algoritmos genéticos, *text mining*, teste de hipóteses, e muito mais. (Turban, et al., 2009).

Muitas empresas do ramo de saúde começaram a usar a análise preditiva para prever os riscos de pessoas adoecerem e para quem os tratamentos serão mais custosos no futuro, de posse de tal informação é possível reduzir custos e atuar de forma preventiva. (Turban, et al., 2009).

De igual forma a análise preditiva aplica-se no campo das finanças possibilitando determinar por meio de suas técnicas (regressão linear, árvore de decisão, *naive bayes*, *k-nn*, redes neurais<sup>3</sup> e outras) a possível arrecadação dos tributos de sua competência e despesas, levando em conta as diversas variáveis (crescimento econômico, emprego de tecnologia, recursos humanos, entrada e saída de empresas, benefícios fiscais concedidos, inadimplência, entre outras) que envolvem previsão de receitas e despesas que torna complexo de fazer sem o emprego das técnicas de análise preditiva.

## **14 USANDO AS TÉCNICAS DE *MACHINE LEARNING* PARA PREVER INADIMPLÊNCIA DE TRIBUTOS E SONEGAÇÃO FISCAL.**

---

<sup>3</sup> As redes neurais representam uma tecnologia que tem raízes em muitas disciplinas: neurociência, matemática, estatística, física, ciência da computação e engenharia, tendo aplicação em campos tão diversos, como modelagem, análise de séries temporais, reconhecimento de padrões, processamento de sinais e controle, em virtude de uma importante propriedade: a habilidade de aprender a partir de dados de entrada com ou sem professor. (Haykin, 2001).

Uma das questões importantes que atormentam o gestor público na administração tributária é do fato de não poder avaliar com precisão o nível de inadimplência dos tributos, alguns tributos possuem o caráter de sazonalidade e outros são mais diretamente afetados pela crise econômica.

Esta é, portanto, uma questão perturbadora para os gestores públicos, pelo fato de que as despesas são previstas no orçamento levando em conta a projeção da arrecadação tributária, e quando ocorre a frustração de receitas pelos mais variados fatores, o orçamento necessita de ajuste para se adequar à nova realidade.

Principalmente num cenário de crise, esta é uma parte que deve ser bem gerenciada, ou pode levar sérios problemas de caixa.

Com a aplicação das técnicas aprendizagem de máquina (adoção de modelo simples regressão logística ou modelos mais complexos de árvores de decisão) é possível com base no histórico da arrecadação por contribuintes identificar aqueles que podem deixar de pagar, apenas observando o comportamento e as características de seus perfis.

Tendo esta informação, o gestor pode ajustar seu risco, implementar ações e focar os seus esforços para evitar a inadimplência por parte do contribuinte.

Há muitas aplicações no campo prático dos algoritmos de aprendizagem de máquina, caberá ao cientista de dados ou outro profissional que conheça essas técnicas a escolha adequada, conforme dados que ele tenha disponível em qualidade para aplicação dos modelos preditivos.

Exemplo didático traz uma aplicação do algoritmo árvore de decisão construído para prever se, a partir da qualidade do serviço prestado em determinada ocasião, o restaurante terá lucro ou prejuízo. (Silva, et al., 2016, p. 108).

Um outro exemplo didático vem da aplicação da regressão linear ou não-linear para prever se as vendas irão aumentar com a inserção de propagandas, sob o prisma da arrecadação pode ser aplicado considerando a variável aumento de alíquota, implementação de novos mecanismos de combate à evasão verificar o quanto irá impactar a arrecadação. (Silva, et al., 2016).



A literatura apresenta muitos trabalhos em mineração de dados aplicados no suporte à gestão tributária, combate à evasão fiscal.

Destaca-se estudo apresentado por (Souza, 2002) aplicou algoritmo de mapas auto organizáveis das redes neurais artificiais, para classificar potenciais contribuintes sonegadores de ICMS, na Sefaz-GO, a partir da análise do banco de dados da extinta Declaração Periódica de Informações – DPI, enviada mensalmente pelos contribuintes e, atualmente, substituída pela Escrituração Fiscal Digital – EFD.

A dissertação realizada por (Andrade, 2009) que utilizou algoritmos das redes neurais artificiais para agrupamento, seleção de atributos e classificação de contribuinte potenciais sonegadores de ICMS, na Sefaz-BA.

Aplicação do método árvore de decisão com o objetivo de classificar os contribuintes de ISS nas categorias regular e irregular, a partir da análise do banco de dados da Secretaria de Finanças do município de Goiânia, com um índice de acertos de 92,03%. (Levergger, 2013).

Aplicação de técnicas de mineração em um programa de concessão de benefícios ao consumidor conhecido como Nota Legal buscando a melhoria da gestão do programa: a fidelidade das pessoas físicas ao programa e a obtenção de créditos de consumo pelos beneficiários. (Vieira, 2014).

## **15 CONSIDERAÇÕES FINAIS**

Adentramos na era em que as obrigações tributárias das empresas e das pessoas físicas são dispostas ao fisco em formato digital, somado ao volume de informações provenientes dos sistemas transacionais e sistemas externos, o volume de dados a serem manipulados aumentam exponencialmente.

Para tanto requer o emprego de tecnologias e ferramentas adequadas ao tratamento desses dados a fim de gerar conhecimento que possibilite ao fisco o combate de forma eficaz à evasão fiscal.

Se por um lado a informatização trouxe ganhos de produtividade, economicidade, por outro, trouxe a complexidade de lidar com esses dados sem perda de qualidade, com segurança e com objetividade.

Sem ferramentas adequadas é inviável e impraticável pensar em auditar com agilidade as informações tributárias enviadas ao fisco, e mesmo promover cruzamentos ou confronto a fim de verificar se há divergências nas informações.

Nesse contexto os investimentos em tecnologia da informação e comunicação são o caminho único possível para atender os ditames constitucionais que regem a administração pública, dentre os quais destaca-se a eficiência.

Obter eficiência nos procedimentos e controles da arrecadação tributária requer a aplicação de novas ferramentas capazes de lidar com esses grandes volumes de dados mantendo a segurança da informação e possibilitando ao fisco manipular essas informações e efetuar os cruzamentos necessários ao alcance do resultado almejado que é melhorar a arrecadação das receitas tributárias, se antecipando as ações de fraudes e sonegação e combatendo-as com maior rapidez e com maior eficácia.

Os cruzamentos de informações e o emprego de inteligência fiscal são essenciais ao fortalecimento das atividades de auditoria e denotam a importância do trabalho do auditor fiscal e a necessidade preeminente de capacitação técnica para lidar com a nova realidade da fiscalização tributária, a auditoria em arquivos eletrônicos e emprego de técnicas de mineração de dados, de aprendizagem de máquina, assim como, de agentes inteligentes, que otimizam as tarefas repetitivas fortalecem a atividade tributária aperfeiçoando os mecanismos de combate à evasão fiscal.

## **REFERENCIAS BIBLIOGRÁFICAS**

Andrade, H. d. S., 2009. *Um Processo de Mineração de Dados Aplicado ao Combate a Sonegação Fiscal do ICMS*. [Online] [Acesso em 25 08 2016].

Barbieri, C., 2011. *BI2 - Business Intelligence: Modelagem & Qualidade*. Rio de Janeiro: Elsevier.

Brognoli, A., 2010. Implementação de um Sistema de Inteligência de Negócios (BI). 24 04.

Castro, L. N. d. & Ferrari, D. G., 2016. *Mineração de dados*. 1ª ed. São Paulo: Saraiva.

Cleto, N. & Oliveira, F. R., 2010. *Manual de Autenticação de Livros Digitais*. 1ª ed. s.l.:Fiscosoft.

Coêlho, S. C. N., 2003. *Teoria Geral do Tributo, Da Interpretação e Da Exoneração Tributária*. 3ª ed. São Paulo: Dialética.

Coelho, S. C. N., 2010. *Teoria da Evasão e da Elisão em Matéria Tributária*. [Online]  
[Acesso em 07 09 2016].

Figueiredo, G. H. d. B., 2008. Um Novo Paradigma na Auditoria em Meio Digital. *Revista da Receita Federal*.

Grus, J., 2016. *Data Science do Zero*. 1ª ed. Rio de Janeiro: Alta Books.

Haykin, S., 2001. *Redes Neurais*. 2ª ed. Porto Alegre: Bookman.

Leony, M. d. G. S., 2006. Informação como instrumento de inteligência na área fiscal: estudo de caso na Sefaz. BA.

Levergger, T. P., 2013. Mineração de Dados Aplicada à Classificação dos Contribuintes do ISS. *Pontifícia Universidade Católica de Goiás*.

Madeira, R. d. O. C., 2015. Aplicação de Técnicas de Mineração de Texto na Detecção de Discrepâncias em Documentos Fiscais. *EMAP Dissertações*, 29 09.

Mayer, V. & Cukier, K., 2013. *Big Data - Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. 1ª ed. Rio de Janeiro: Elsevier.

Oliveira, F. N. d., 2011. Estratégias para aperfeiçoar o processo de recuperação de receitas tributárias no Estado da Bahia:Um modelo para o ICMS baseado em redes neurais artificiais.

Piccirilli, T. L., 2013. Mineração de Dados Aplicada a Classificação dos Contribuintes do ISS.

Primak, F. V., 2008. *Decisões com B.I (Business Intelligence)*. s.l.:Ciência Moderna.

Provost, F. & Fawcett, T., 2016. *Data Science para Negócios*. 1ª ed. Rio de Janeiro: Alta Books.

Queiroga, R. M., 2005. Uso de técnicas de data mining para detecção de fraudes em energia elétrica. *Biblioteca Central da Universidade Federal do Espírito Santo*.

Russel, S. & Norvig, P., 2013. *Inteligência Artificial*. Rio de Janeiro: Elsevier.

Serra, L., 2002. *A essência do Business Intelligence*. São Paulo: Berkeley.

Silva, L. A. d., Peres, S. M. & Boscaroli, C., 2016. *Introdução à Mineração de Dados com Aplicações em R*. 1ª ed. Rio de Janeiro: Elsevier.

Souza, A. J. V. d., 2002. O Uso de Mapas Auto-Organizáveis para Classificar Contribuintes do ICMS.

Turban, E., Sharda, R., Aronson, J. E. & King, D., 2009. *Business Intelligence: Um Enfoque Gerencial para a Inteligência do Negócio*. Porto Alegre: Bookman.

Vieira, M. H. P., 2014. Aplicação de técnicas de mineração em um programa de concessão de benefícios ao consumidor: o caso do Programa Nota Legal do Distrito Federal.