

Aprimorando Conferências Tributárias Através da Computação Cognitiva Utilizando a Plataforma Watson

Leonardo Fabro da Rocha¹

¹Instituto de Ciências Exatas e Geociências – Universidade de Passo Fundo (UPF)
CEP 99052-900 – Passo Fundo– RS – Brasil

leonardo@sollucao.net

Resumo. *Com a obrigatoriedade da emissão de notas fiscais no formato eletrônico, a quantidade de informações geradas a partir destes documentos vem aumentando consideravelmente nos últimos anos. Do mesmo modo, os escritórios de contabilidade estão aprimorando as conferências tributárias sobre esses dados gerados, com o intuito de evitar a escrituração de informações incorretas perante a Receita Estadual. Por tudo isso, o objetivo do presente trabalho é a utilização de uma técnica de inteligência artificial para verificar se a alíquota do ICMS está emitida de maneira correta em diversos itens de uma NF-e. A fundamentação teórica destaca a carga tributária brasileira, o imposto sobre circulação de mercadorias, os sistemas de informação aplicados à contabilidade, a inteligência artificial aplicada a contabilidade, a aprendizagem de máquina, as redes neurais artificiais e os trabalhos relacionados ao tema. A metodologia da pesquisa enfatiza a utilização da plataforma Watson com uma integração via webservice para realizar a predição dos dados desejados. Os resultados aferidos evidenciam que a utilização de uma técnica de inteligência artificial para o problema proposto foi eficaz, notando-se uma redução no tempo da análise da alíquota dos itens das notas fiscais, sendo que o analista fiscal analisou somente os itens que poderiam estar incorretos, além de ter indicado itens com a tributação realmente incorreta. A conclusão reitera que o objetivo proposto foi atingido com base nos resultados apresentados.*

Abstract. *With the obligatory emission of electronic receipts, the amount of information generated to leave of those documents increased considerably in the last years. In the same way, the accounting offices are perfecting the tax conferences on those generated data, with the intention of avoiding the bookkeeping of incorrect information before the State Revenue. For all this, the objective of the present work is the use of a technique of artificial intelligence to verify the rate of ICMS is emitted in a correct way in several items of a NF-e. Theoretical basis emphasizes the Brazilian tax burden, the tax on goods circulation, information systems applied to accounting, artificial intelligence applied to accounting, machine learning, artificial neural networks and related work. The research methodology emphasizes the use of the Watson platform with a webservice integration to predict the desired data.*

The checked results evidence that the use of a technique of artificial intelligence for the proposed problem was effective, being noticed a reduction in the time of the analysis of the rate of the items of the notes, being that the fiscal analyst analyzed only the items that could be incorrect, besides having really indicated items with the taxation incorrect. The conclusion reiterates that the proposed objective was reached with base in the presented results.

1. Introdução

Com o avanço da era digital nos processos fiscais e a obrigatoriedade de empresas emitirem notas fiscais de forma eletrônica (NF-e e NFC-e), a quantidade de informações geradas de maneira digital aumentou consideravelmente nos últimos anos para os escritórios contábeis. Segundo dados do site oficial da NF-e, já foram emitidas 19,377 bilhões de notas no formato digital e existem atualmente 1,537 milhões de emissores ativos. [1]

Esse modelo de NF-e foi criado pela Receita Federal, sendo um subprojeto do SPED (Sistema Público de Escrituração Digital da Receita Federal), com o objetivo de facilitar e aumentar a fiscalização das informações prestadas nesse documento. Nesse modelo implementado, foi exigido um nível maior de detalhamento dessas informações, a qual possibilita uma fiscalização mais efetiva e ágil por parte dos órgãos públicos.

Ao mesmo tempo, a legislação brasileira é de uma complexidade considerável, sendo que para ser definido a tributação de um produto, deve ser observado “n” fatores que podem influenciá-la. Com essa complexidade, ficamos nas mãos dos sistemas corporativos das empresas e de seus usuários, que podem configurá-los de maneira incorreta e emitir notas com tributações equivocadas, gerando grandes prejuízos às suas empresas, quando as mesmas forem submetidas a uma fiscalização dos órgãos públicos.

Existem várias técnicas computacionais que podem solucionar parcialmente esse problema, sendo que é necessário realizar na maioria dos sistemas computacionais diversas configurações para que possa ser possível verificar se os impostos de uma NF-e estão corretos. Porém, aplicando técnicas de IA, é possível solucionar o problema de maneira mais prática e eficaz, com mínima configuração e intervenção humana.

Deste modo seria possível realizar o processo da validação das alíquotas de ICMS de maneira mais ágil e rápida, reduzindo o risco de algum erro humano não ser identificado, possibilitando a localização do mesmo a tempo para a correta escrituração e utilizando o tempo ganho com essa automatização para realizar tarefas que agreguem mais valor ao cliente - como o estudo da legislação estadual para identificar possíveis vantagens fiscais sobre os produtos comercializados pelo mesmo, entre outras possibilidades. Essa ferramenta torna-se um grande diferencial tecnológico no mercado contábil para o escritório Solução Contabilidade - parceiro do presente trabalho - que possui como um de seus objetivos a constante evolução tecnológica para agilizar seus processos internos, melhorando cada vez mais a qualidade dos mesmos.

Portanto o objetivo do presente trabalho é demonstrar a utilização de uma técnica de inteligência artificial para verificar se a alíquota do ICMS (Imposto Sobre Circulação de Mercadorias) está sendo emitida de maneira correta em diversos ítems de uma NF-e, sendo que o analista fiscal da empresa contábil que irá realizar os testes do

presente trabalho (Solução Contabilidade) deverá analisar somente as notas que apresentarem divergências, ao invés de realizar a conferência de todas as notas, uma a uma.

2. Revisão Bibliográfica

Nesta seção serão descritos os tópicos relevantes para a pesquisa deste artigo e os trabalhos relacionados com o tema.

2.1. Carga Tributária Brasileira

Segundo uma recente pesquisa do Banco Mundial, 68,4% (sessenta e oito vírgula quatro por cento) dos lucros das corporações nacionais são destinados ao pagamento de impostos [9], tornando o Brasil um dos países que possuem a maior carga tributária ao redor do mundo. [8] Nossa carga tributária é dividida em impostos sobre a propriedade (Ex. IPVA e IPTU), sobre a renda (Ex. IRPF) e sobre produtos e serviços (Ex. ICMS e IPI).

Os impostos sobre produtos e serviços são divididos em duas categorias: federais e estaduais. Os impostos federais (como por exemplo PIS e COFINS) são recolhidos para a máquina pública federal e suas regras, na maioria das vezes, valem sobre todos os estados da união. Entretanto, possuímos uma peculiaridade maior nos impostos estaduais (como por exemplo ICMS), pois cada unidade federativa pode possuir a sua própria legislação e suas próprias regras tributárias, divergindo na maioria das vezes, umas das outras.

As empresas brasileiras, com essa alta carga tributária e com a alta complexidade desses impostos, são as mais afetadas. Geralmente as mesmas costumam praticar preços elevados aos consumidores finais para conseguir sanar todas as dívidas tributárias com a máquina pública. Para complementar, a complexidade das regras tributárias atuais eleva os riscos das empresas ficarem expostas a riscos de autuações pelo poder público, por informações tributárias prestadas de maneira incorreta, na maioria das vezes oriundas de erros de configuração de sistemas ou de informação digitada pelo usuário incorretamente.

2.2. Imposto Sobre Circulação de Mercadorias (ICMS)

O Imposto Sobre Circulação de mercadorias é um imposto não-cumulativo, sendo que a cada fase da cadeia econômica de circulação o contribuinte tem o direito de compensar o mesmo imposto pago nas etapas anteriores, também se tornando mais complicado pelo fato de ser cobrado por unidades subnacionais, com alíquotas diferenciadas e com concessão de subsídios fiscais em cada unidade que, por muitas vezes não obedecem a Constituição Federal (CF/88). [2]

Existem diversos fatores que podem influenciar na definição da tributação correta do ICMS, sendo que esses fatores podem variar entre as unidades federativas. Na maioria das vezes, o CFOP (Código Fiscal de Operações e Prestações), aliado ao estado de origem e destino da NF-e, podem definir a alíquota correta a ser aplicada. Entretanto, cada estado define algumas NCM(s) (Nomenclatura Comum do Mercosul) que terão algum benefício fiscal ou alíquota majorada em determinadas operações. Também podemos ter alguns benefícios fiscais aplicados somente a alguns produtos específicos dentro de cada NCM.

Diante de toda essa complexidade, é natural que as empresas possuam dificuldades na hora da emissão de uma NF-e, pois torna-se difícil averiguar toda a legislação vigente para a definição da alíquota de ICMS correta a ser utilizada.

2.3 Sistemas de Informações Aplicados à Contabilidade e Apuração de Imposto

Dentro do mercado contábil existem diversas ferramentas e sistemas que realizam a apuração de impostos e demonstram a contabilidade das empresas que contratam os escritórios contábeis. No mercado nacional se destacam algumas desenvolvedoras de software, como Questor¹ e Domínio². A Questor é uma desenvolvedora de softwares contábeis que iniciou suas atividades em 1985 e não demorou muito tempo para iniciar com o desenvolvimento de softwares próprios. Hoje em dia possui todos os programas necessários para a escrita contábil (contábil, fiscal e pessoal) e já está começando a investir em alguns softwares para auditoria de dados fiscais, como o analista fiscal digital e analista fiscal contábil. [10] A Domínio foi fundada em 1998 e possui somente módulos para a realização da parte fiscal e contábil. Em 2012 foi adquirida pela norte-americana Thomson Reuters e se tornou um dos principais softwares contábeis. [11]

Entretanto, a Sollução Contabilidade (empresa contábil que irá realizar os testes do trabalho) possui softwares internos, por acreditar que esses sistemas garantem o diferencial de seu negócio. Um desses módulos, denominado AUDITA, possui diversas conferências fiscais e cruzamentos que são realizados em cima dos arquivos XML das NFes emitidas por seus clientes, sendo que os erros encontrados pelo mesmo são corrigidos e os clientes são avisados sobre todos esses erros, para realizarem a correção em seu sistema. Esse módulo, entretanto, não possui a capacidade computacional de realizar a validação proposta no presente trabalho. Por essa razão, está aprimorando seu sistema e buscando novas formas que estão emergindo para realizar essas validações.

Uma delas é a Inteligência Artificial, pois a empresa acredita que esse é o futuro do seu mercado e nela visualiza que vários processos podem ser otimizados com a utilização dessa nova tecnologia.

2.4 Inteligência Artificial Aplicada à Contabilidade

A Inteligência Artificial dentro dos escritórios contábeis ainda é um assunto muito pouco difundido. Quando falamos desse assunto surgem inúmeras dúvidas: O uso dessa tecnologia irá substituir o profissional? Como essa tecnologia pode ser usada no dia a dia? Quais os benefícios de usá-la? O receio de ser substituído por um software de gestão pode amedrontar o contador e trazer receios sobre o uso dessa tecnologia. [5]

Entretanto, o dia a dia do contador 2.0 muda com essa tecnologia e se torna consultivo e segmentado. Com mais tempo para se concentrar em aspectos de gestão, esse profissional entrega consultoria de valor por meio de informações valiosas que mantêm a empresa do cliente em uma crescente constante, lidera o escritório com planejamento e objetivo para encontrar soluções para futuros problemas e passa a se

¹ Questor é um software de serviços contábeis e fiscais desenvolvido pela empresa Questor - Sistemas Inteligentes.

² Domínio é um software de serviços contábeis e fiscais desenvolvido pela empresa Thomson Reuters.

dedicar a encontrar formas de criar oportunidades de negócios, para adquirir novos clientes. [5]

Também destaca-se que várias empresas contábeis serão extintas pois, com a otimização dos processos de escrituração contábil e fiscal, somente os melhores e mais capacitados profissionais continuarão no mercado. Também é importante ressaltar que é necessário realizar a transformação para o contador 2.0 o quanto antes, para possuir mais chances de continuar “vivo” na profissão.

2.5 Aprendizagem de Máquina

O aprendizado de máquina é uma subárea da Inteligência Artificial, que realiza a aprendizagem através da experiência, conforme as tarefas são executadas. Sendo assim, a mesma aprende a melhor maneira de se resolver um problema e consegue estruturar o conhecimento existente. [4]

Existem vários métodos de aprendizado de máquina, sendo que podemos citar o aprendizado por hábito, que tem como característica o programa aprender por experiência de acordo com o que foi informado anteriormente e o método do aprendizado por conceito, o qual analisa todas as hipóteses e demonstra qual é a correta[4].

Nesse método temos a “hipótese mais geral”, o que significa que se não existe nenhuma possibilidade correta, o sistema achará a mais próxima do correto. Entretanto estes métodos possuem alguns problemas pois, nem sempre o usuário quer saber a hipótese correta e sim a mais comum.[4]

Existem diversos algoritmos de aprendizagem de máquina, cujo intuito é permitir que, após um determinado treinamento com um certo conjunto de dados cujas instâncias têm classificação conhecida, uma máquina seja capaz de interpretar novos dados e classificá-los de maneira apropriada a partir de uma generalização do que lhe foi apresentado anteriormente. Alguns algoritmos têm como inspiração os sistemas biológicos (como as RNAs e os Algoritmos Genéticos), os processos cognitivos (Raciocínio Baseado em Casos), o aprendizado simbólico (Árvores de Decisão) e as teorias estatísticas (SVMs). Três são os paradigmas que indicam a forma de aprendizado do algoritmo: ([12])

1. Supervisionado: utiliza supervisores para obter o melhor modelo desejado na fase de treinamento.
2. Não-supervisionado: permite que o algoritmo aprenda a representar ou agrupar as entradas segundo alguma medida de similaridade entre as mesmas.
3. Por reforço: o aprendizado ocorre por meio de recompensas, dependendo do desempenho apresentado.

2.6 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são sistemas de inteligência artificial que formam e aplicam o conhecimento prévio para novos problemas ou situações inéditas. O mesmo foi proposto por McCulloch e Pitts (1943) e surgiram como uma tentativa de simulação matemática do princípio básico de funcionamento dos componentes mais elementares do sistema nervoso humano. Uma RNA é formada por uma camada de neurônios de entrada (sensorial), outra de saída (transmissora) e pode ter uma ou mais

camadas intermediárias ou ocultas (FIGURA 1) que podem ser necessárias para a representação de tipos mais complexos de informação. [2]

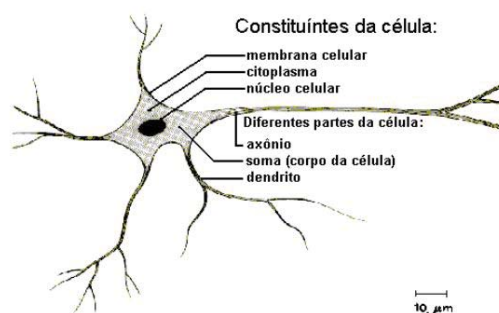


Figura 1 -Desenho de um neurônio biológico

Na maior parte dos modelos de RNAs, com exceção da camada de entrada, as camadas são constituídas de nodos, unidades de processamento cujo funcionamento baseia-se nos princípios biofísicos dos neurônios biológicos (sinapses). Cada nodo possui terminais de entradas, que recebem os sinais transmitidos e, para cada conexão associada a esses terminais, existe um valor chamado peso, o qual representa a intensidade de sua respectiva força sináptica. [2]

É através da função de ativação, também chamada de função restritiva (HAYKIN, 2001), que a saída do nodo é calculada. Existem diversos tipos de funções de ativação e o seu uso está associado ao tipo de problema a ser tratado, sobretudo quanto ao tipo e ao intervalo de saída que o nodo deve produzir. A saída de um nodo pode ser transmitida para outros nodos, como uma de suas entradas ou, caso o nodo esteja na camada de saída, será uma das saídas da rede [2]

2.7 Trabalhos Relacionados

O processo de seleção dos estudos primários utilizados nesta seção é descrito a seguir: os estudos foram analisados e avaliados a partir de seus títulos e resumos, a fim de satisfazerem o critério de elegibilidade. Foram utilizados dois grupos de palavras chaves:

- Grupo A: Inteligência Artificial;
- Grupo B: Contabilidade, Administração, Tributação e ICMS;

A string de busca utilizado foi "A1 + (B1 OR B2 OR B3 OR B4)." e as bases que foram realizadas as pesquisas foram: Google Academic, Science Direct, CAPES e DEVMEDIA. Foram considerados somente os artigos com no máximo 5 anos de publicação, escritos em Inglês ou Português. Os estudos selecionados nessa primeira fase foram analisados mais detalhadamente e, a partir da leitura dos métodos, conclusões e de seu conteúdo, foram elencados dois trabalhos para esta seção.

2.7.1 SIRC

O SIRC (Sistema de Identificação de Risco de Contribuintes) é uma rede neural de modelo perceptron de múltiplas camadas, do tipo alimentada adiante (feedforward), completamente conectada, composta por duas camadas ocultas e treinada com o algoritmo de retropropagação de erro (Backpropagation Error). O objetivo da mesma é verificar se um contribuinte é passível ou não de uma fiscalização estadual.[2]

A entrada da rede é composta por 36 campos numéricos, cada um deles associado a uma característica que compõe o perfil de um contribuinte. Cada entrada é fornecida por -1 (não possui) ou 1 (possui) tal característica. No desenvolvimento do software pelo profissional que implementa a rede neural artificial SIRC, foi adotada a linguagem Java™ da Sun Microsystems®. Sobre os resultados da rede, a SIRC se mostrou eficaz na seleção de contribuintes, centrando o foco naqueles que realmente podem vir a ser passíveis de fiscalização. [2]

2.7.2 CLASSIFICAÇÃO DOS CONTRIBUINTES DE ICMS DA SEFAZ-GO

Neste trabalho são realizadas técnicas de mineração de dados, onde são analisados grandes volumes de dados em busca de correlações e padrões relevantes para os usuários tomadores de decisão. Normalmente, as técnicas de mineração de dados executam as tarefas de classificação e agrupamento dos dados e descoberta de regras de associação. [3]

A classificação é uma das mais populares tarefas de mineração de dados, sendo que é implementada por um algoritmo de aprendizagem de máquina que, a partir de um conjunto de dados de treinamento previamente classificados, relaciona-os de acordo com suas características, produzindo um modelo que é validado a partir de um conjunto de dados teste. [3]

A árvore de decisão é representada graficamente como uma árvore, com nós e ramos, mas no sentido invertido, sendo que cada nó contém um teste e seus resultados vão formando os demais ramos. [3]

Sendo assim, o estudo 2 utilizou o software WEKA, implementando o algoritmo J48 e analisando uma massa de dados de contribuintes de ICMS do estado de Goiás que sofreram infrações entre 2013 a 2016. Foram utilizados em alguns cenários os dados do contribuinte, registro de apuração do ICMS, registro do inventário e resultado de auditoria de ICMS. Sendo assim, com os novos dados, em seu melhor cenário, a taxa de classificação correta ficou em torno de 84%. [3]

3. Metodologia

Nesta seção será descrito como o trabalho e as pesquisas foram realizadas para realizar a implementação do sistema em questão. A seguir, será exibido um fluxo do trabalho, para melhor visualização dos tópicos que serão tratados nessa seção.

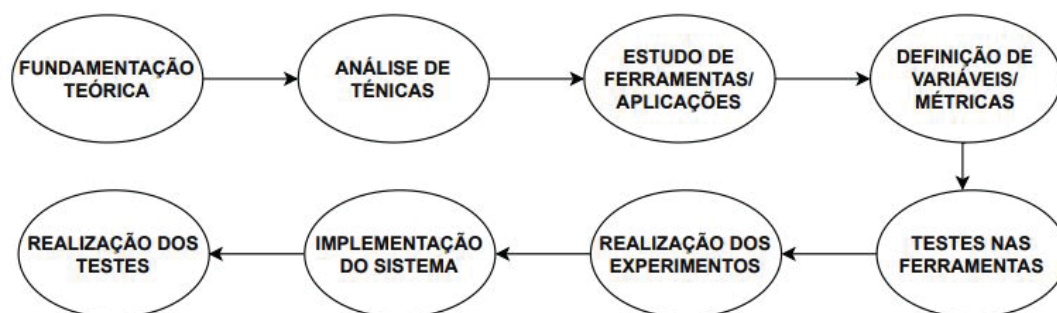


Figura 2 -Fluxo do desenvolvimento do trabalho

Primeiramente foi analisado o cenário atual e foi constatado que o problema principal é identificar quando um cliente do escritório contábil emite uma NFE com

tributação incorreta de ICMS, devido ao grande volume de NFEs que o escritório recebe e a complexidade da legislação atual de ICMS, que no paradigma da programação tradicional não se tem solução viável. Um exemplo seria um cliente emitir uma NFE com tributação normal - no estado do RS seria 18% - porém o ítem dessa nota possui uma redução na alíquota, pois em determinada operação o ítem pertence a cesta básica brasileira.

Com o problema definido, foi realizada a busca por trabalhos relacionados ao tema. Nesse ponto foi identificada uma grande carência de trabalhos que tratam do tema em questão no âmbito da inteligência artificial. Mesmo assim, foram encontrados dois trabalhos que se aproximam do assunto em questão, que foram descritos no tópico “2.7” do presente trabalho.

Tendo em vista algumas das tecnologias utilizadas nos trabalhos relacionados, foi realizada uma pesquisa sobre técnicas de inteligência artificial para verificar qual se adaptaria melhor ao problema exposto anteriormente. A técnica estudada foi “Redes Neurais”, a qual foi explicada no capítulo 2, item 2.6.

Sendo assim, foram realizados testes em ferramentas com essa técnica, por se entender que ela poderia solucionar o problema. Primeiramente, foi utilizada a ferramenta WEKA, que é um pacote de software em Java. O mesmo tem como objetivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na subárea da Inteligência Artificial, dedicada ao estudo de aprendizagem de máquina. [6]

A segunda ferramenta utilizada foi o NeuroSolutions, que é uma extensão para o Excel. O mesmo permite criar de maneira muito simples uma Rede Neural utilizando uma planilha com dados normalizados para realizar o treinamento da rede e a posterior validação. [7]

A terceira ferramenta testada foi o IBM Watson, o qual é uma plataforma de desenvolvimento de aplicações de Inteligência Artificial, com enfoque em Machine Learning e com diversos serviços para utilizá-la da melhor maneira possível. Um deles é o IBM Studio, o qual é uma plataforma para realizar treinamentos com modelos personalizados e com dados próprios. [3]

Para começar, é necessário criar um serviço do IBM Studio com as configurações básicas para utilizá-lo. Após a criação, ao entrar na plataforma, veremos que o mesmo é orientado a projetos, sendo necessário criar um para utilizar as funcionalidades da plataforma. Ao criá-lo e acessá-lo, vemos que existem diversas abas e funcionalidades para serem exploradas, sendo que iremos focar somente no que é de interesse do presente artigo.

Existem três técnicas de treinamento disponíveis:

1. Binary classification: Classifica os dados em duas categorias (verdadeiro ou falso). Exemplo: Classificar se o participante possui risco de determinadas doenças.
2. Multiclass classification: Classifica os dados em diversas categorias. Exemplo: Classificar um consumidor em um determinado estilo musical.
3. Regression: Realiza a predição do valor através de um conjunto contínuo de valores.

Para realizar os testes nas ferramentas mencionadas, foi inicialmente realizado a normalização dos dados disponíveis no banco de dados da empresa contábil parceira, com todos os itens das NF-e's emitidas pelos seus clientes desde o início do ano de 2018. Foram elencadas algumas variáveis que poderiam identificar a maioria das situações tributárias do ICMS:

- CFOP: Código Fiscal de Operação;
- NCM: Nomenclatura Comum do Mercosul;
- Estado de Origem da Mercadoria;
- Estado de Destino da Mercadoria;
- Origem do ICMS. Ex: Nacional, Estrangeira;
- CST de ICMS: Código da Situação Tributária;
- Percentual de ICMS.EX: 18%;

A seguir, é exibido um exemplo dos dados normalizados:

```
CFOP, NCM, POS_IES, ORIGEM, DESTINO, ORI_CST, CST_ICM, PERCENTUAL
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
6910, 33029090, 1, 43, 15, 0, 00, 7.00
```

Figura 3 - Normalização dos dados

O primeiro teste foi realizado na ferramenta WEKA, com as técnicas de “Logistic Regression” e “MultilayerPerceptron” (Redes Neurais), sendo que nenhuma das duas técnicas conseguiu “aprender” com os dados fornecidos e a predição não foi realizada de forma correta. O segundo teste realizado foi na ferramenta NeuroSolutions, que também apresentou o mesmo problema relatado anteriormente. Porém, na realização do terceiro teste com a ferramenta IBM Watson, utilizando o IBM Studio e as três técnicas mencionadas anteriormente, foi verificado um aumento das taxas assertivas para o problema em questão. Por essa razão, foi decidido utilizá-lo para a realização do presente trabalho.

Para realizar o treinamento de uma rede, o Watson utiliza alguns serviços para tal, como o “Machine Learning”. O mesmo possibilita os computadores a “aprenderem” com algoritmos específicos, facilitando a criação de modelos de treinamento para uma determinada rede e ajudando na visualização de comparações entre algoritmos de aprendizagem, para facilitar a decisão de qual utilizá-lo de maneira mais ágil.

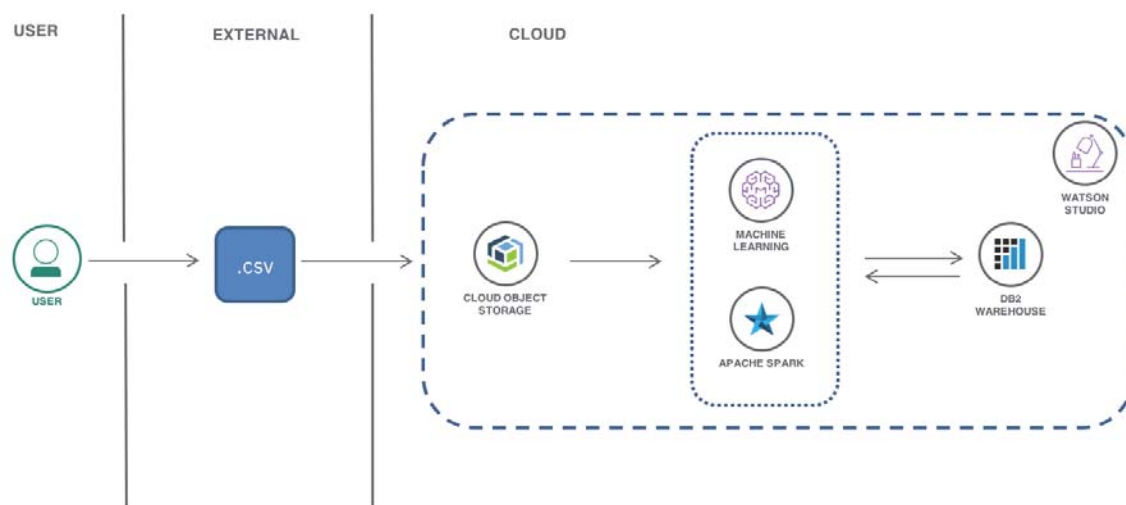


Figura 4 -Panorama do funcionamento da plataforma Watson Machine Learning. Fonte: <https://console.bluemix.net/docs/tutorials/create-deploy-retrain-machine-learning-model.html#build-deploy-test-and-retrain-a-predictive-machine-learning-model>

Nos dados utilizados no presente trabalho, foi realizada a comparação da taxa assertiva de todos os algoritmos de aprendizagem de máquina disponíveis para treinamento, os quais serão apresentados a seguir - com a exceção do tipo "LogisticRegression", o qual possui uma forma diferente de cálculo de assertividade dos demais tipos de algoritmo.

ALGORITMO	TIPO	RESULTADO	ASSERTIVIDADE
DecisionTreeClassifier	MultiClass Classification	Excelente	0,98450
RandomForestClassifier	MultiClass Classification	Excelente	0,98990
NaiveBayes	MultiClass Classification	Razoável	0,71403
LogisticRegression	Binary Classification	Ruim	0,50812
DecisionTreeClassifier	Binary Classification	Péssimo	0,44470
RandomForestClassifier	Binary Classification	Péssimo	0,25481
GBTClassifier	Binary Classification	Péssimo	0,20187

Tabela 1 - Tabela de resultados dos testes realizados com os algoritmos

Após a análise da comparação entre as taxas, foi escolhida a técnica de "Multiclass classification" com o algoritmo "Random Forest Classifier", que possuiu a maior taxa assertiva - próxima a 100% - entre todos os resultados obtidos. O mesmo

constrói diversas árvores de decisão para produzir o rótulo que é um modo de cada árvore de decisão. [14]

Após a definição da técnica, foi realizado alguns experimentos com as variáveis do problema. Foi realizada a tentativa de adicionar o CNPJ do destinatário e o código de barras dos itens da NF-e para verificar a taxa assertiva com o algoritmo selecionado. Entretanto, a taxa caiu para 90% com a primeira variável e para 70% com a segunda, sendo necessário retirá-los para não comprometer a taxa da rede.

Após a definição final das variáveis e de realizar o treinamento da rede, foi necessário realizar uma integração com os dados do escritório contábil parceiro para que a rede pudesse fornecer as respostas (alíquotas) correspondentes para as novas NF-es que os clientes emitissem, sendo que o objetivo é comparar a resposta com a alíquota informada na nota. Para tal, o IBM Studio disponibiliza um WEB Service - com padrão JSON - para cada modelo criado na plataforma, sendo possível enviar os dados para análise da rede e obter as respostas da mesma em uma aplicação local.

Para isso, foi criado um cliente na linguagem Java, utilizando a plataforma de desenvolvimento NetBeans, para consumir o WEB Service mencionado anteriormente. Foi realizada uma consulta SQL para a obtenção dos dados, por estabelecimento (CNPJ) e mês/ano desejado e, após isso foi percorrida essa tabela e montado uma string em formato JSON para o envio. O formato de envio do arquivo é o seguinte:

```
{"fields": ["CFOP", "NCM", "ORIGEM", "DESTINO", "ORI_CST", "CST"], "values": [CFOP_1, NCM_1, ORIGEM_1, DESTINO_1, ORI_CST_1, CST_1], [CFOP_2, NCM_2, ORIGEM_2, DESTINO_2, ORI_CST_2, CST_2]}
```

Também é necessário possuir credenciais específicas para o presente processo, sendo que foi realizada a criação das mesmas para utilização da plataforma. Após o envio, o tempo de retorno varia dependendo da quantidade de “linhas” que foram enviadas para análise. Após o retorno com as respostas - no formato JSON - o sistema realiza a verificação de cada linha, comparando se o conteúdo retornado do campo “PERCENTUAL” é igual ao conteúdo do campo “PERCENTUAL” da NF-e. Se for, o sistema grava em uma lista auxiliar para cada linha analisada a seguinte mensagem: "CERTO! VALOR NOTA E WATSON: X". Caso contrário, será gravado: "ERRO: VALOR NOTA: X | VALOR WATSON: Y".

É importante ressaltar que os dados enviados retornam na mesma ordem do envio e, para realizá-lo, o sistema gera o conteúdo no formato “JSON” para a comunicação com o Watson. O sistema grava os dados enviados ao Web Service e também os retornados do mesmo em duas listas distintas para, após um determinado tratamento do retorno, o sistema realizar a comparação de cada linha das duas listas, para verificar se o percentual das duas posições é idêntico e atestar se o Watson localizou alguma divergência.

Para facilitar a realização dos testes e da validação, foi implementada uma tela em um ambiente WEB já existente, no módulo denominado “fiscal”. O mesmo utiliza tecnologias como ANGULAR 5 e BOOTSTRAP, sendo que toda a parte de controle de autenticação e comunicação do frontend com o backend já está funcional. Através da integração dos dados do escritório contábil com o serviço do Watson, que foi realizada anteriormente, levantou-se uma URL em um WEB Service REST - também já existente

- para consumi-la. Os parâmetros necessários para a URL são a unidade e o mês/ano das notas desejadas.

A seguir, segue imagens da tela que foi utilizada para os testes:



Figura 5 - Tela para testes sem dados

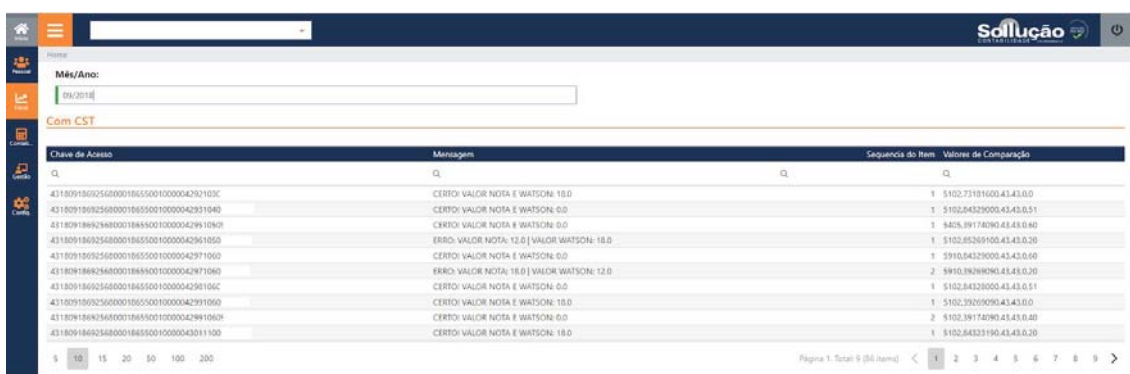


Figura 6 - Tela para testes com dados

É importante frisar que o frontend (imagem acima) realiza a requisição ao serviço de integração dos dados fiscais com o Watson e que esse processo será realizado mensalmente por um analista fiscal do escritório contábil.

Após o retorno dos dados, o analista fiscal deverá realizar um filtro para analisar os registros que possuem a mensagem que contém a palavra “ERRO”, para verificar se os erros que a validação encontrou são verídicos. Isso quer dizer que ele deverá analisar as notas com os itens que retornarem inconsistentes, verificando a legislação vigente e atestando se o sistema encontrou de fato uma inconsistência verdadeira. Os resultados desse processo serão descritos no ítem 4 do presente artigo.

4. Análise e Discussão de Resultados

Nesta seção serão apresentados os resultados obtidos nos testes realizados no presente trabalho e a análise dos mesmos.

Na realização do treinamento da rede no Watson, foi utilizado um conjunto de dados com 476 mil registros, com itens de notas fiscais emitidas de janeiro a agosto de 2018, sendo que 60% dos dados foram utilizados para treinamento, 20% para testes e 20% para Holdout. Com isso, a rede conseguiu uma assertividade de 98,99%, o que é um percentual bastante expressivo e significativo para treinamentos.

Na fase de testes e de validação da rede, foram selecionadas 76 empresas, a qual possuíam o regime de tributação “Lucro Presumido” ou “Lucro Real”, que realizaram a

emissão de NF-es no mês de setembro de 2018, para serem submetidas a avaliação do Watson. A quantidade de notas emitidas no respectivo mês pelas empresas selecionadas foi de 19.703, com 67.644 ítems. Após a análise dos mesmos, foram localizados 144 possíveis erros que foram revisados pelo analista fiscal.

Para isso, o analista fiscal analisou somente a tributação de cada ítem que tinha a possibilidade de estar incorreto, ao invés de analisar todos os ítems do período. Nesta análise, localizou 17 itens que estavam com a alíquota informada incorretamente na nota fiscal, o que acarretaria problemas fiscais a empresa e ao escritório contábil junto a Receita Estadual.

Com os resultados obtidos anteriormente, foram realizadas as seguintes constatações:

- A rede não conseguiu criar o padrão da alíquota em alguns casos através da NCM. Exemplo: Produtos de cosmética com alíquota de 25% que a rede entendia que era 18%;
- A rede não conseguiu distinguir uma alíquota diferenciada, no caso da mesma possuir a sua determinação através da descrição do produto. Exemplo: NCM com alíquota de 18%, que quando possuir a descrição “Veículo para Transporte de Mercadoria” sua alíquota é reduzida para 12%;
- A CST em alguns casos é um campo que pode influenciar fortemente a formação do padrão para a alíquota e, quando informado incorretamente na emissão da nota, pode levar o Watson a realizar uma classificação de uma alíquota idêntica a da nota, mesmo que o CFOP e a NCM indiquem que a alíquota seja distinta. Exemplo: CST 40 com alíquota zero, quando a tributação correta do produto é alíquota de 18% com CST 00;
- A rede conseguiu localizar alíquotas que estavam informadas incorretamente, através do estado de origem e destino. Exemplo: Alíquota informada de 12%, sendo que o correto é 18%;
- A rede conseguiu localizar alíquotas incorretas que não existem em alguns casos. Exemplo: Alíquota informada de 40% ou 5% na nota;
- A rede conseguiu localizar alíquotas informadas incorretamente para uma determinada NCM em determinados CFOPs. Exemplo: Alíquota de 18%, sendo que o correto é 0%, e vice-versa;
- Existiu uma situação em que o Watson e a nota estavam com o percentual incorreto, sendo que o mesmo informou uma alíquota de 17% e ela estava com 18%, sendo que a alíquota correta era 12%.
- A rede localizou 17 erros em ítems que a alíquota estava incorreta na nota, sendo que a empresa foi avisada para realizar a correção no seu sistema e evitar futuros problemas com fiscalizações do governo estadual.
- Notou-se que o analista fiscal focou a sua análise em somente 144 registros - menos de 1% do total de itens - tendo reduzido seu trabalho significativamente, possuindo assim oportunidade de focar o tempo ganho em outras atividades que agreguem mais valor ao cliente - como análises tributárias ou estudos de legislações.

Com base nos resultados obtidos, o uso da Inteligência Artificial atrelado a área tributária para a análise da alíquota de ICMS se mostrou bastante eficaz no presente trabalho. Ainda que o uso da computação tradicional seja uma alternativa, o uso da IA

se mostrou muito mais prático e rápido de se implementar, não exigindo parametrizações por parte do usuário e sendo mais ágil para realizar a análise dos itens se comparado ao modelo tradicional, por utilizar os dados das notas fiscais emitidas pelos clientes para gerar inferências do resultado final. Com isso, notou-se uma significativa redução no tempo do processo de análise da alíquota de ICMS dos itens das notas fiscais emitidas pelos clientes do escritório contábil, possibilitando que o analista fiscal realizasse somente a análise de possíveis erros desses itens, tendo também indicado alguns itens com a tributação incorreta na nota, possibilitando que o cliente evite futuros problemas com a Receita Estadual.

5. Considerações Finais

Com o avanço da tecnologia sobre os processos fiscais - especialmente na área tributária - a quantidade de informações geradas diariamente vem aumentando de maneira considerável a cada ano. Aliado a esse avanço a legislação brasileira é bastante complexa, sendo que a tributação de um produto envolve diversas variáveis que podem influenciar na sua correta tributação. Nesse cenário, o escritório contábil fica dependente da emissão correta de notas pelos seus clientes que em algumas situações podem emitir notas com tributações incorretas por desconhecimento da legislação estadual.

Ainda que existam diversas técnicas computacionais que podem solucionar parcialmente o problema, estas demandam grande parametrização por parte do usuário para realizar a correta validação da alíquota presente na nota. Entretanto, com a aplicação de técnicas de Inteligência Artificial é possível verificar se a alíquota de ICMS está emitida de maneira correta em uma NF-e de maneira mais prática e fácil, como foi demonstrado no capítulo anterior, sendo que o analista fiscal teve que analisar somente os itens que apresentaram divergência, ao invés de realizar a conferência sobre todos os itens um a um.

Os resultados aferidos evidenciam que a utilização de uma técnica de inteligência artificial para a resolução do problema proposto foi eficaz, sendo que o mesmo reduziu o número de itens para a análise do analista fiscal e conseguiu localizar erros de tributação em documentos fiscais eletrônicos que estavam com a sua alíquota de ICMS informada incorretamente no mesmo. Com isso podemos identificar a grande vantagem em se utilizar aprendizagem de máquina ao invés do uso de regras para solucionar o problema proposto, pois a mesma utiliza os próprios dados das notas fiscais emitidas pelos clientes para gerar inferências do resultado final, ao invés de serem criadas inúmeras regras pelo analista fiscal para tentar resolver o problema.

Com isso, o presente artigo traz uma grande contribuição no âmbito da Inteligência Artificial para a área tributária, pois demonstra que é possível realizar processos complexos e que demandam grande tempo para serem executados de maneira mais rápida e confiável, possibilitando que os analistas fiscais dos escritórios contábeis realizem tarefas que agreguem mais valor para o cliente, como o estudo da legislação estadual para identificar possíveis vantagens fiscais sobre os produtos vendidos, entre outras possibilidades.

6. Referências

- [1] Portal da Nota Fiscal Eletrônica, Informações Sobre Estatísticas. Disponível <<http://www.nfe.fazenda.gov.br/portal/infoEstatisticas.aspx>>. Acessado em: 14 de junho de 2018.
- [2] NOBRE DE OLIVEIRA, Francisco. ESTRATÉGIAS PARA APERFEIÇOAR O PROCESSO DE RECUPERAÇÃO DE RECEITAS TRIBUTÁRIAS NO ESTADO DA BAHIA: UM MODELO PARA O ICMS BASEADO EM REDES NEURAIS ARTIFICIAIS. Universidade Federal da Bahia, 2011. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/7435/1/OLIVEIRA%2c%2520FN.pdf>>. Acessado em: 20 de junho de 2018.
- [3] MEIRELES ROCHA, Santiago. MINERAÇÃO DE DADOS APLICADA À CLASSIFICAÇÃO DOS CONTRIBUINTES DE ICMS DA SEFAZ-GO. Pucgoias.edu, 2017. Disponível em: <<http://tede2.pucgoias.edu.br:8080/bitstream/tede/3892/2/SANTIAGO%20MEIRELES%20ROCHA.pdf>>. Acessado em: 20 de junho de 2018.
- [4] MACHADO DA SILVA, Brigiane; VANDERLINDE, Marcos. INTELIGÊNCIA ARTIFICIAL, APRENDIZADO DE MÁQUINA. Universidade do Estado de Santa Catarina, 2012. Disponível em: <http://www.ceavi.udesc.br/arquivos/id_submenu/387/brigiane_machado_da_silva_marcos_vanderlinde.pdf>. Acessado em: 20 de junho de 2018.
- [5] RIGO DA CRUZ JACOBSEN, Gabriel. REDES BAYESIANAS. Ceavi, 2012. Disponível em: <http://www.ceavi.udesc.br/arquivos/id_submenu/387/gabriel_rigo_da_cruz_jacobsn.pdf>. Acessado em: 25 de junho de 2018.
- [6] Wikipedia, Weka. Disponível em: <<https://pt.wikipedia.org/wiki/Weka>>. Acessado em: 10 de julho de 2018.
- [7] NeuroSolutions, Características. Disponível em: <<https://pt.wikipedia.org/wiki/Weka>>. Acessado em: 10 de julho de 2018.
- [8] ROCHA BATISTA, Natan. O Brasil possui a maior carga tributária do mundo?. JusBrasil, 2015. Disponível em: <<https://natanrocha.jusbrasil.com.br/artigos/239372506/o-brasil-possui-a-maior-carga-tributaria-do-mundo>>. Acessado em: 18 de junho de 2018.
- [9] Impostos no Brasil: por que a carga tributária brasileira é tão alta?. Blog de Valor, 2018. Disponível em: <<https://andrebona.com.br/impostos-no-brasil-por-que-a-carga-tributaria-brasileira-e-tao-alta/>>. Acessado em: 19 de junho de 2018.
- [10] Questor, Soluções. Disponível em: <<http://questor.com.br/>>. Acessado em: 10 de julho de 2018.

- [11] Thomson Reuters, Visão Geral Domínio. Disponível em:
<<http://www.dominiosistemas.com.br/visao-geral/>>. Acessado em: 10 de julho de 2018.
- [12] VALERIO NETTO, R. et al. TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA ANÁLISE DE IMAGENS OFTALMOLÓGICAS. Universidade de São Paulo, 2003. Disponível em
<<http://www.lbd.dcc.ufmg.br/colecoes/wim/2003/003.pdf>>. Acessado em: 10 de agosto de 2018.
- [13] IBM, O que é o Watson?. Disponível em:
<<https://www.ibm.com/watson/about>>. Acessado em: 06 de setembro de 2018.
- [14] IBM, Visão Geral dos Construtores de Modelos. Disponível em:
<<https://dataplatfom.cloud.ibm.com/docs/content/analyze-data/ml-model-builder.html?audience=wdp>>. Acessado em: 10 de outubro de 2018.