

Reformulação de Decisões de Órgãos Fazendários

Reform of Treasury Bodies Decisions

Helena Cristina de Albuquerque Bastos^{1,2,3}  orcid.org/0000-0002-9849-9561

Rômulo César Dias de Andrade^{1,3}  orcid.org/0000-0002-0561-7507

Marcelo Pita^{2,4}  orcid.org/0000-0001-7582-4651

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

² Serviço Federal de Processamento de Dados (SERPRO)

³ Centro de Informática, Universidade Federal de Pernambuco, Recife, Brasil

⁴ Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil

E-mail do autor principal: helena-cristina.bastos@serpro.gov.br

RESUMO

O objetivo deste trabalho foi fazer uma mineração de dados, nos órgãos fazendários de qualquer esfera de governo, de forma a verificar se existem padrões nas bases de dados que levam a reformulação do reconhecimento do direito creditório do contribuinte, em uma instância superior. Para atingir este objetivo foi realizada a modelagem do processo Restituição, Compensação e Ressarcimento do Contribuinte, na notação BPMN, obtendo-se o modelo To Be. Dessa forma, se adquiriu o conhecimento necessário do negócio e, então, foi realizada uma análise das informações com base nas técnicas descritivas e preditivas de mineração dos dados deste processo. Este artigo relata a experiência realizada para atingir o objetivo supracitado e os resultados obtidos. Tais resultados se mostraram de alto valor descritivo e preditivo, com regras que indicam relacionamentos importantes entre as variáveis de interesse e a capacidade de antecipar, com alta precisão, a aceitação ou não de manifestações de inconformidade.

PALAVRAS-CHAVE: Modelagem de Processo de Negócio (BPM); Data Lake; Mineração de Dados;

ABSTRACT

The aim of this paper is identifying in the Government Treasury Agencies of any Government Area, which factors lead to the reformulation of the recognition of the taxpayer's credit right, in a higher instance. To reach that goal, the process modeling Restitution, Compensation and Reparation of the taxpayer, in the BPMN notation. Thus, the necessary knowledge is gained and so the information analysis is done based on the data mining of this process. This article reports on the experience gained to achieve the aforementioned objective and the results obtained. These results were shown to be highly descriptive and predictive, with rules that indicate important relationships between the variables of interest and the ability to anticipate, with high precision, the acceptance or non-compliance of manifestations of nonconformity.

KEY WORDS: Business Process Modeling (BPM); Datalake; Data Mining;

1 INTRODUÇÃO

Considerando uma visão simplificada do ordenamento jurídico-tributário do governo o pagamento de qualquer imposto pode ser realizado de maneira compulsória ou declaratória. Ao contribuinte é disponibilizado o preenchimento de um pedido de restituição (PER) através de um programa gerador de declaração (PGD) com o objetivo de restituir um imposto considerado pago indevidamente. Se o contribuinte estiver em débito com o governo deverá utilizar a declaração de compensação (DCOMP) revertendo o crédito em compensação.

Os pedidos de manifestação de conformidade geram um processo administrativo que demanda recursos preciosos da administração pública o que motiva o emprego de técnicas que melhor permitam decidir pela aceitação ou não de tais pedidos.

Assim, este trabalho tem como objetivo aplicar técnicas de mineração de dados ao processo decisório de forma a levantar padrões que aprimorem o processo de aceitação de processos de restituição fazendária, também prevendo se estes processos serão deferidos pela justiça. Uma abordagem baseada na dinâmica do processo de negócio considerando as atividades de restituição, compensação e ressarcimento ao contribuinte foi realizada partindo da notação BPMN, obtendo-se o modelo *To Be*. Essa abordagem possibilitou o conhecimento necessário do negócio para a aplicação das técnicas de mineração de dados.

Este artigo está organizado da seguinte forma: na Seção 2 apresentamos a fundamentação teórica necessária a essa pesquisa. A Seção 3 apresenta as atividades de mineração de dados realizada a partir da modelagem de processos de negócio. Finalmente, a Seção 4 apresenta as conclusões.

2 REFERENCIAL TEÓRICO

2.1 BPM

O gerenciamento de processos de negócio, difundido sob a denominação de BPM (sigla em inglês para: *Business Process Management*) ganhou corpo principalmente a partir dos anos 2000 sendo bastante influenciado pelos avanços da tecnologia de informação. Ele provê às organizações a possibilidade de incrementar a competitividade e sustentabilidade, em tempos

de incerteza no mercado, globalização crescente e constante mudança nas condições dos negócios segundo Bruin [1].

Conforme mostram Paim et al, a literatura apresenta muitas definições do que se pode entender por gerenciamento de processos de negócio [2]. De acordo com estes autores, esta denominação tornou-se mais popular a partir da segunda metade da década de 1990. Para Jeston e Nellis, BPM não é um conceito simples de definir e menos ainda de implementar [3]. O ciclo de BPM envolve as atividades de identificação, planejamento, modelagem, execução, automação e otimização de processos de negócio. De acordo com a ABPMP, BPM é uma disciplina gerencial com o objetivo de atingir resultados consistentes e alinhados às metas estratégicas da organização, através do gerenciamento de processos voltado para sua identificação, projeto, execução, documentação, medição, monitoramento e melhoria contínua [4]. Podemos encontrar na literatura, definições distintas de sequências de atividades ou de ciclos para a implantação do BPM. Todavia, apesar das diferenças encontradas entre as abordagens sobre o tema, quase a sua totalidade é baseada no ciclo PDCA.

Outra definição útil para BPM é proposta por Weske: um conjunto de conceitos, métodos e técnicas (fortemente fundamentadas no uso de tecnologia da informação) para suportar o desenho, administração, configuração, determinação de uso e análise de processos de negócio [5]. Todavia, vale ressaltar que o uso do termo BPM também é, frequentemente, encontrado na literatura, como modelagem de processos de negócio (*business process modeling*) segundo Recker et al., 2006 [6]. De acordo com Bruin, evidências incidentais sugerem que o uso inconsistente da terminologia no domínio de BPM torna difícil o entendimento e comparação de pesquisas nesta área [1]. Para fins deste artigo, adotou-se a visão menos usada atualmente, entendendo o termo como *modeling* (modelagem) em vez *management* (gerenciamento) que é a mais utilizada.

Embora possa ser compreendida primariamente como uma abordagem de gerenciamento, BPM apresenta-se como um campo de interesse multidisciplinar. Particularmente, como campo de aplicação de tecnologia da informação nas organizações, BPM tem se destacado como uma abordagem

poderosa para inserção de TI, gerando uma vantagem competitiva no negócio central das instituições. BPMS (*Business Process Management Systems*, ou em português: sistemas de gerenciamento de processos de negócio) é um método frequentemente usado nas organizações, que facilitam a implementação em *software* dos princípios de BPM. Ferramentas de BPMS costumam contemplar em suas funcionalidades, recursos como: diagramação e documentação de processos; execução automatizada do controle do fluxo de trabalho (*workflow*); possibilidade de criação de indicadores gerenciais de processos em painéis de controle; armazenamento e trâmite de documentos eletrônicos com possibilidade de certificação digital; integração com sistemas legados através de filosofia SOA (*service oriented architecture*).

A representação dos processos é peça fundamental no desenvolvimento das diversas fases do ciclo BPM, desde as etapas de desenho e modelagem até no auxílio ao controle e refinamento. Alguns padrões de representação dos processos são utilizados em BPM, como UML, IDEF e EPC, todavia, a notação BPMN (*Business Process Modeling Notation*) tem se tornado efetivamente o padrão mais utilizado para modelagem de processos, contando atualmente com inúmeras ferramentas compatíveis com as suas especificações. A notação BPMN é mantido pelo OMG (*Object Management Group*) [7], e tem o objetivo de fornecer uma notação facilmente compreensível por todos os usuários de negócio, desde os analistas de negócio, os desenvolvedores de tecnologia de suporte à execução dos processos e os gestores de processos, eliminando a lacuna entre o desenho dos processos de negócios e a sua implementação.

2.2 Mineração de Dados

A mineração de dados envolve obrigatoriamente o conhecimento do negócio, técnicas adequadas para cada tipo de problema e infraestrutura de TIC (*hardware, software* e rede). No processo de mineração de dados, a extração, transformação e carga (ETL) de múltiplas fontes torna-se necessária. Dados brutos em geral passam por rotinas de tratamento, ou pré-processamento, tais como: limpeza de dados, redução de dimensionalidade, normalização e discretização

de atributos.

A mineração de dados pode ser analisada a partir da sua capacidade em resolver determinados grupos de problemas, tais como:

- Descrição: os dados utilizados em uma análise podem descrever um comportamento ou tendência;
- Classificação: consiste em determinar classes dos dados através de modelos que exploram o poder discriminativo das características para cada classe;
- Regressão: similar o processo de classificação, porém busca prever um valor no domínio contínuo de um registro a partir de um modelo gerado através de dados conhecidos;
- Agrupamento: na tarefa de agrupamento de registros similares são identificados. Cada grupo (cluster) é formado por um conjunto de registros similares entre si e idealmente diferentes dos registros pertencentes aos demais grupos;
- Associação: a tarefa de associação consiste em identificar padrões de co-ocorrência nos valores dos atributos analisados. Em geral, a associação é expressa através de regras do tipo Se X então Y; em que, X e Y são conjuntos de atributos categóricos.

Os problemas de mineração de dados, dividem-se em dois grandes grupos: atividades preditivas e atividades descritivas. No contexto deste trabalho, focaremos na atividade preditiva, através de classificação, e na atividade descritiva, com descoberta de regras de associação, conforme está detalhado na Seção 3.

Na classificação, o objetivo é prever classes ou rótulos para dados desconhecidos. Recebe-se como entrada o valor correto de uma função desconhecida para entradas específicas e tenta-se inferir esta função, de acordo com Zaki e Meira Jr. [8]. Existem diversas famílias de técnicas para classificação: KNN (k-Nearest Neighbors), Naïve Bayes, Árvores de decisão, Redes Neurais, SVM (*Support Vector Machine*), dentre outras.

A tarefa de descoberta de regras de associação consiste em encontrar padrões frequentes, associações, correlações, ou estruturas causais que ocorram

simultaneamente de forma frequente em um conjunto de dados, segundo Zaki e Meira Jr [8]. Existem diversos algoritmos para extração de regras de associação: Apriori, FP-Tree (*Frequent Pattern Tree*), *Rapid Association Rule Mining* (RARM), *Special Issues of Association Rule Mining*, *Multiple Concept Level ARM*, *Multiple Dimensional ARM*, *Constraints based ARM*, *Maintaining of Association Rules*, dentre outros. Suporte e confiança são usados como filtros, para diminuir o número de regras, gerando apenas regras de maior qualidade. Considerando regras do tipo X então Y, podemos definir suporte (s) como a fração das transações que contêm X e Y, e confiança (c) como a frequência com que Y aparece nas transações que contêm X.

Outro conceito importante neste artigo é o de *data lake* (Lago de Dados) que são dados armazenados, integrados e acessados em seu formato nativo e utilizados para agregar valor a diferentes serviços e soluções. Tornar disponível todos os dados do cliente, de forma a permitir o armazenamento, tratamento, análise e disponibilização de maneira fácil e em grande quantidade. O *data lake* é a estrutura sobre a qual estão apoiadas as soluções de *big data*. O *big data* não gera valor por si. A geração de valor é quando conseguimos criar *insights*, em coleções de dados, que resultam benefícios tangíveis para o cliente segundo Marr [10]. Os desafios e objetivos na montagem desta arquitetura foram: obter tecnologia para *big data* (*hardware* e *software*); dados atômicos e rapidamente disponíveis (carga diária ou próxima ao tempo real); não descartar dados após período de tempo; *self-service* BI

(flexibilidade, agilidade e autonomia dos usuários); possibilitar o acesso dos sistemas transacionais e atendimento das diversas capacidades analíticas. O conteúdo do *data lake* é composto de tabelas desnormalizadas, dados de eventos, dados históricos, dados mestres, dados higienizados, regras de negócio aplicadas e dados brutos.

3 Mineração dos dados com base em BPM

O processo de negócio para a atividade de restituição, compensação e ressarcimento do contribuinte trata do pagamento de qualquer imposto que pode ser realizado de maneira compulsória ou declaratória. Ao contribuinte, é disponibilizado o preenchimento, utilizando-se de um programa gerador de declaração (PGD), de um pedido de restituição (PER) com o objetivo de restituição financeira de um imposto considerado pago indevidamente. Alternativamente, a declaração de compensação (DCOMP) permite a utilização de um crédito revertendo-o em compensações a débitos existentes junto ao governo.

Considerando o processo de negócio nas atividades de restituição, compensação e ressarcimento do contribuinte, foi realizada a modelagem deste processo junto com todos os envolvidos (cliente e equipe de TI), obtendo-se o modelo TO BE do processo na ferramenta BPMS denominada ORIX, conforme apresentado de forma macro na Figura 1 e descrito de forma detalhada a seguir.

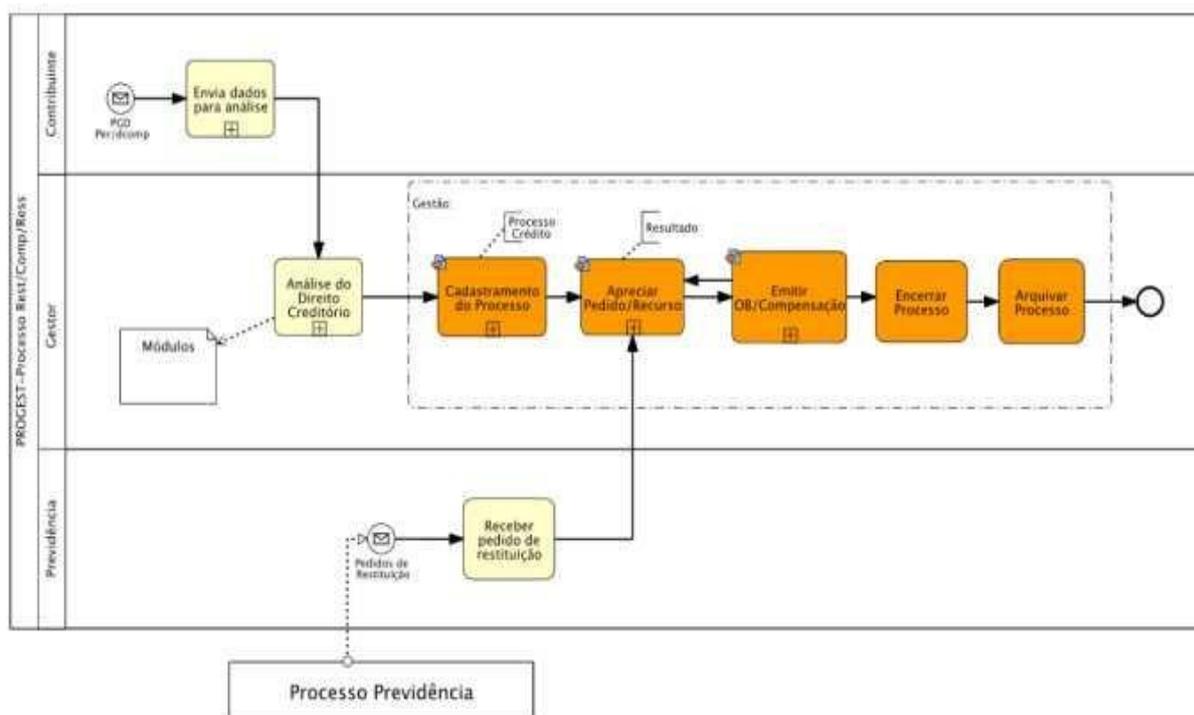


Figura 1: Modelagem de processo de negócio Restituição, Compensação e Ressarcimento do Contribuinte.

Uma vez enviadas as diversas declarações, há um processamento realizando as compensações entre os débitos e créditos, procedimento denominado Reconhecimento do Direito Creditório (RDC), que como resultado final emite um documento, denominado Despacho Decisório, a ser entregue ao contribuinte contendo uma relação com os valores fiscalmente reconhecidos e um resultando com um saldo a pagar ou a receber.

Caso discorde do cálculo realizado pelo órgão fazendário, no âmbito administrativo, permite-se ao contribuinte uma manifestação de inconformidade (defesa do contribuinte), dirigida, inicialmente, ao órgão juridicamente competente, considerado um primeiro recurso de decisão. Em caso de indeferimento da manifestação de inconformidade, parcial ou total, cabe ao contribuinte recorrer a uma última instância administrativa.

No intuito de evitar o trabalho relativo aos procedimentos decorrentes da aceitação de uma manifestação de inconformidade, foi percebida a necessidade de identificação de padrões de comportamento nas instâncias de recurso que refletem em um acatamento do pleito do contribuinte.

Com base no conhecimento adquirido na modelagem do processo de negócio supracitado, identificamos algumas questões relevantes para o cliente, tal como identificar se existem regras e padrões nos dados que levam a reformulação do reconhecimento do direito creditório do contribuinte, em uma instância superior. Inicialmente, foi realizada uma extração de dados do órgão fazendário sem possibilitar a mínima identificação individualizada do contribuinte, conforme solicitado pelo cliente.

Assim, foram disponibilizados dados genéricos, envolvendo características associadas a perfis e comportamentos dos contribuintes com os campos valores dos créditos solicitados e manifestações de inconformidades aceitas (campo valor do recurso e valor do resultado) pelos órgãos institucionais de julgamento do órgão fazendário.

Para as análises, optou-se por uma classificação dos campo tipo de crédito, localidade do contribuinte e atividade econômica, somente na condição em que os órgãos deferiram o pedido do contribuinte, em detrimento do Despacho Decisório. Tornou-se

necessário um processo inicial de tratamento dos dados disponibilizados pelo órgão fazendário e sua importação para o ambiente da ferramenta Knime.

Em seguida, identificou-se a necessidade de descoberta de conjuntos frequentes que fazem com que uma manifestação de inconformidade seja acatada. Assim, por solicitação do cliente e na tentativa de identificação de regras de associação, elaborou-se o fluxo de trabalho na ferramenta Knime, conforme apresentado na Figura 2.

O processo inicia-se com um tratamento do arquivo recebido, eliminando-se colunas com informações não pertinentes ao problema, assim como colunas com valores sem utilidade ("Não se Aplica", "N/A") em todas as linhas. Em seguida, selecionam-se as colunas que servirão de base para análise dos resultados (tipo de crédito, localidade da empresa detentora do crédito e grupo/atividade econômica da empresa declarante, valor da PERDCOMP e valor resultado da manifestação de inconformidade). Finalmente, no componente de descoberta de regras, os valores de suporte e confiança mínimos foram, respectivamente, 1% e 80%.

Durante o processo de geração de regras, vários valores foram configurados para determinar um número considerado aceitável para as regras geradas pela ferramenta. Baseando-se no problema apresentado, a consequência mais destacada seriam os valores de inconformidade, ou seja, em quais situações as inconformidades vêm sendo aceitas e quais os intervalos de valores.

Observa-se, em uma análise simplificada de somente uma das linhas de resultado, que as maiores manifestações de inconformidade, do universo de dados disponibilizados, pertencem as "Atividades de sociedades de participação", tipo de crédito com valores de PERDCOMP ultrapassando R\$ 2.000.000,00 (dois milhões de reais), gerando um valor de inconformidade em um intervalo aproximado entre R\$ 725.000,00 (setecentos e vinte e cinco mil reais e R\$ 197.000.000,00 (cento e noventa e sete milhões de reais).

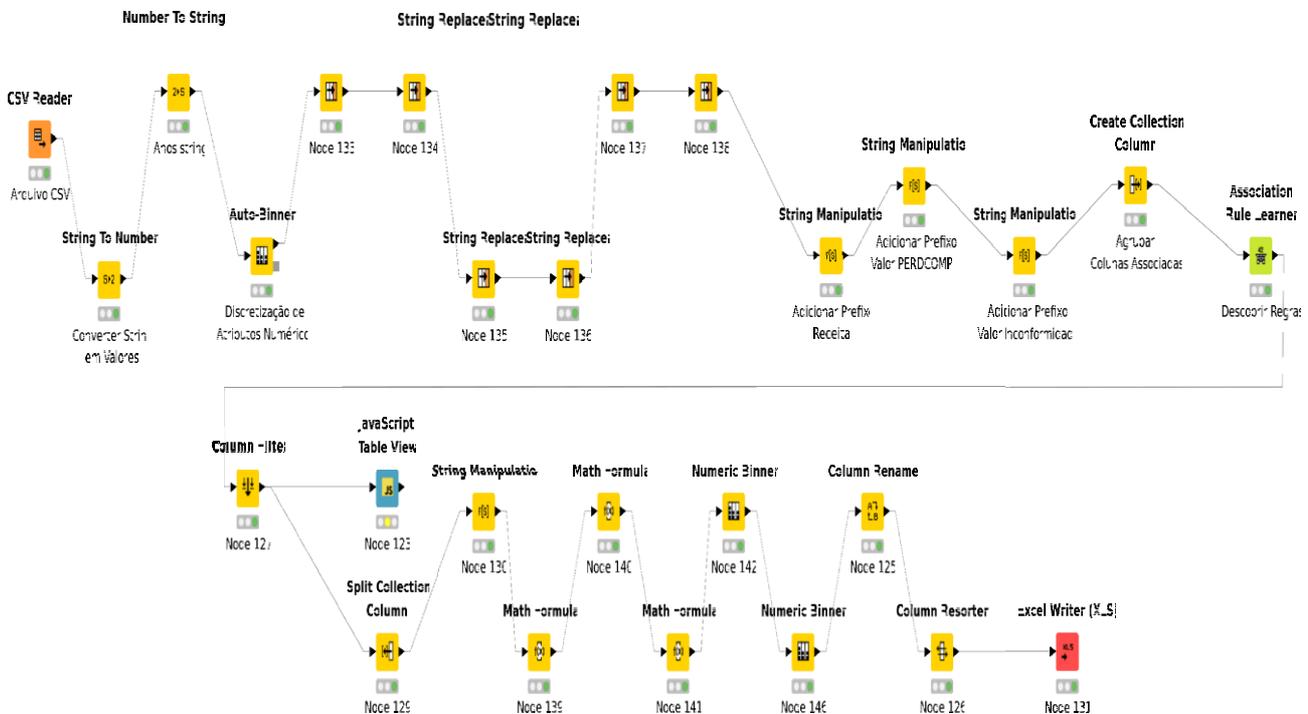


Figura 2: Modelo KNIME para descoberta de regras de associação

Foram realizadas as seguintes extrações de dados:

1. Despachos decisórios com data de julgamento da manifestação de inconformidade (julgados), independente do resultado. Cerca de 1.300 linhas, despachos emitidos em 2016, 2017 e 2018.
2. Despachos decisórios com data de julgados, cerca de 1.300 linhas, despachos emitidos em 2016, 2017 e 2018.

O resultado esperado da modelagem foi identificar regras de não aceitação da manifestação (extração 1) e regras de aceitação da manifestação de inconformidade (extração 2).

Os resultados da modelagem descritiva com descoberta de regras de associação foram analisados pelo cliente para uma efetiva interpretação conclusiva de todo o processamento.

Com base na identificação destes estes padrões, foi possível a implementação de novos critérios nos sistemas transacionais de forma a evitar a emissão de despachos decisórios que levem ao reconhecimento do direito creditório do contribuinte.

Para a modelagem preditiva com classificação foram realizadas as seguintes extrações (treinamento e operação/scoring, respectivamente):

1. Despachos decisórios com data de julgamento da manifestação de inconformidade (julgados), cerca de

1.300 linhas, despachos emitidos em 2016, 2017 e 2018.

2. Despachos decisórios com data de manifestação de inconformidade (julgados) e sem data de julgamento da manifestação de inconformidade (não julgados), cerca de 17.000 linhas, despachos emitidos em 2016, 2017 e 2018.

O resultado esperado da modelagem preditiva foi alcançado, uma vez que se identificou a probabilidade, expressa em grau de certeza, de determinado Despacho Decisório ter aceita a sua manifestação de inconformidade. A construção da solução na ferramenta NIME é apresentada na Figura 3.

Vale ressaltar que neste processo de mineração de dados, a extração, transformação e carga (ETL) de múltiplas fontes foi necessária. Conforme pode ser observado nos modelos apresentados nas Figuras 2 e 3, os dados brutos passaram por rotinas de tratamento e pré-processamento, tais como: limpeza de dados, redução de dimensionalidade, normalização e discretização de atributos.

No contexto deste trabalho, a mineração dos dados foi realizada a partir do data lake cuja arquitetura é apresentada na Figura 4.

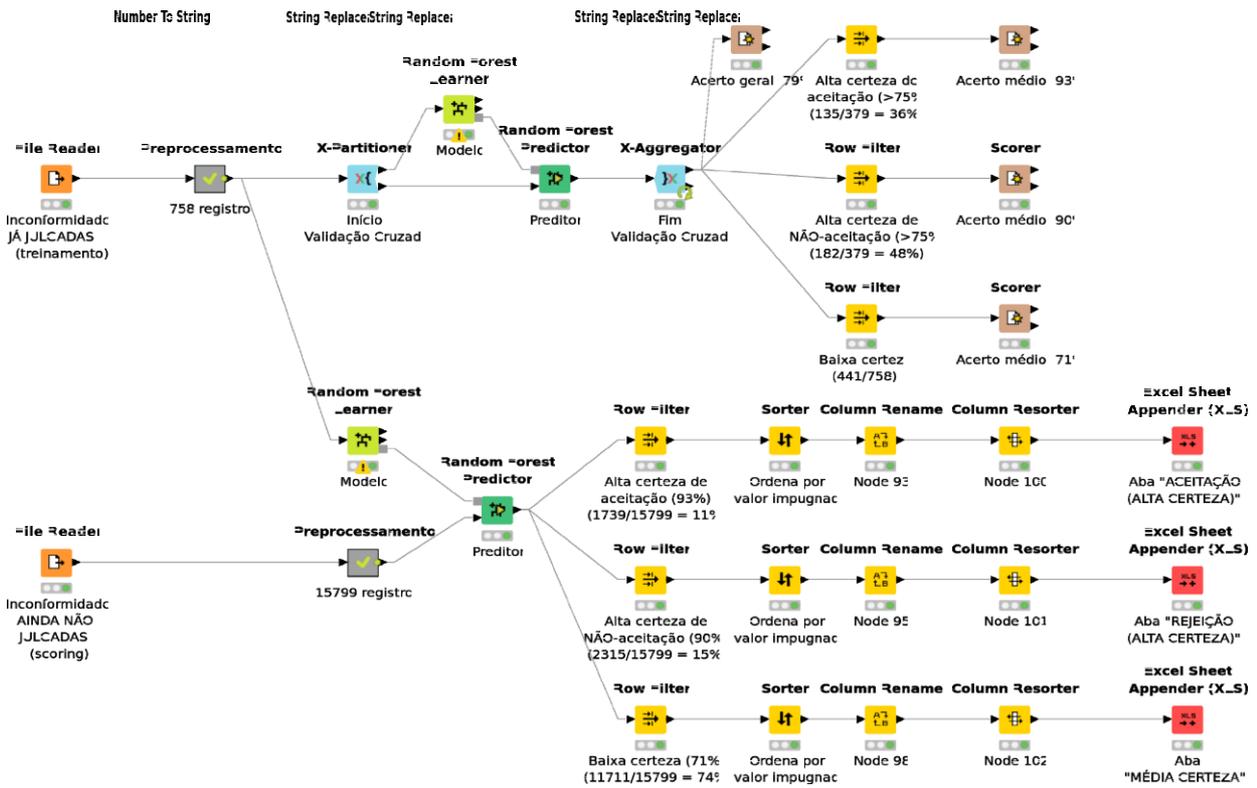


Figura 3: Modelo KNIME para classificação.

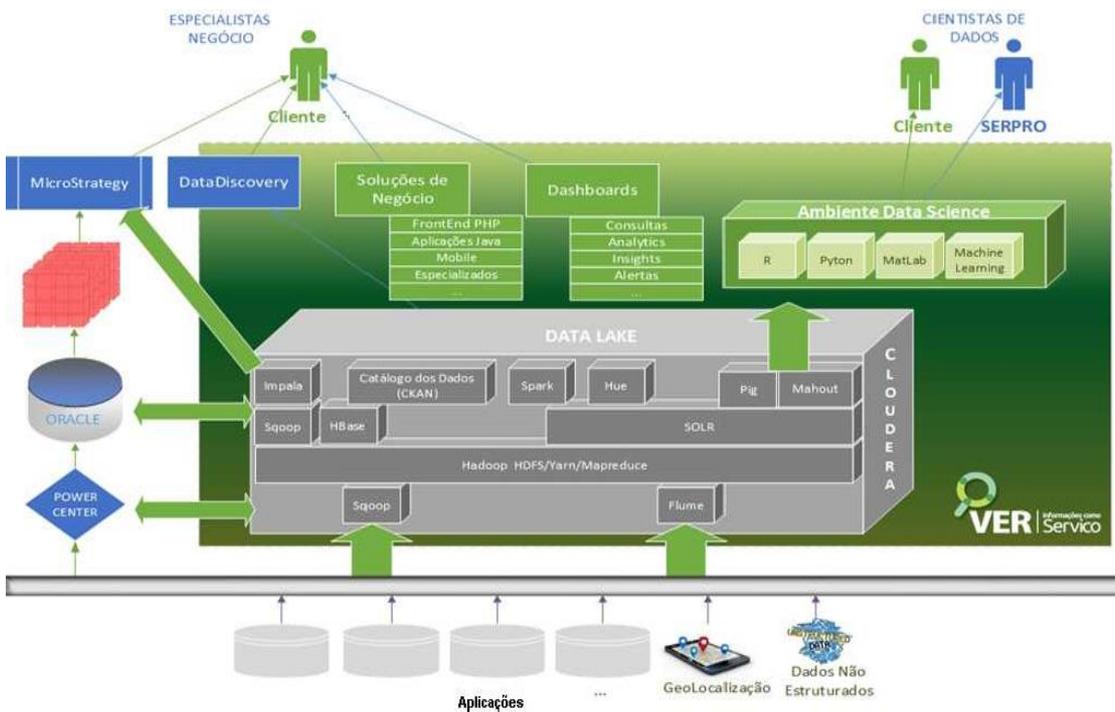


Figura 4: Arquitetura física do Data Lake.

4 CONCLUSÃO

Este trabalho apresentou a aplicação de técnicas de mineração a partir de BPM, onde pode-se verificar a importância do conhecimento do cliente sobre as regras de negócio que foi padronizado através da modelagem de processo.

Verificou-se também a necessidade de uma equipe multidisciplinar com conhecimento do domínio do negócio, além de BPM, estatística, análise de sistema e ciência de dados.

Os resultados alcançados na exploração de técnicas de mineração de dados sobre os processos de negócio (restituição, compensação e ressarcimento do contribuinte) se mostraram de alto valor descritivo e preditivo gerando regras que indicam relacionamentos importantes entre as variáveis de interesse, bem como a capacidade de antecipar, com alta precisão, a aceitação ou não de manifestações de inconformidade

Para trabalhos futuros a área fazendária pode ser vista como um excelente campo para exploração através de estudos empíricos.

REFERÊNCIAS

[1] BRUIN, T. de. Insights into the Evolution of BPM in Organizations. *In: AUSTRALASIAN CONFERENCE ON INFORMATION SYSTEMS*, 18., 2007, Toowooba. **Proceedings**, n. 43., Australia: ACIS, 2007. Disponível em: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1047&context=acis2007>.

[2] PAIM, R. **Gestão de Processos: pensar, agir e aprender**. Porto Alegre: Bookman. 2009.

[3] JESTON, J.; NELIS, J. Business Process Management: practical guidelines to successful implementations, 2006. *In: BALDAM et al. Gerenciamento de processos de negócios. BPM-Business Process Management*. 2. ed. São Paulo: Érica. 2009.

[4] ABPMP. BPM CBOK. Versão 3.0. 2013.

[5] WESKE, M. **Business Process Management. Concepts, Languages, Architectures**. 2. ed. New York: Springer, 2012. DOI: 10.1007 / 978-3-642-28616-2.

[6] RECKER, J. Opportunities and constraints: the current struggle with BPMN. **Business Process Management Journal**, v. 16, n. 1, p. 181-201, 2010.

[7] OMG. Business Process Model and

Notation (BPMN), Version 2.0., c2011.

[8] ZAKI, M., MEIRA-JR., W. **Data Mining and Analysis: Fundamental Concepts and Algorithms**. Cambridge: Cambridge University Press, 2016.

[9] TAURION, C. Big data e o data lake. *In: TI ESPECIALISTA: DESENVOLVENDO IDEIAS*, 2014. Disponível em: <http://www.tiespecialistas.com.br>. Acesso em: 22 nov. 2018.

[10] MARR, B. Why only one of the 5 Vs of big data really matters. *In: IBM BIG DATA & ANALYTICS HUB [Blog]*, 2015. Disponível em: <https://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters>. Acesso: 10 dez. 2018.