

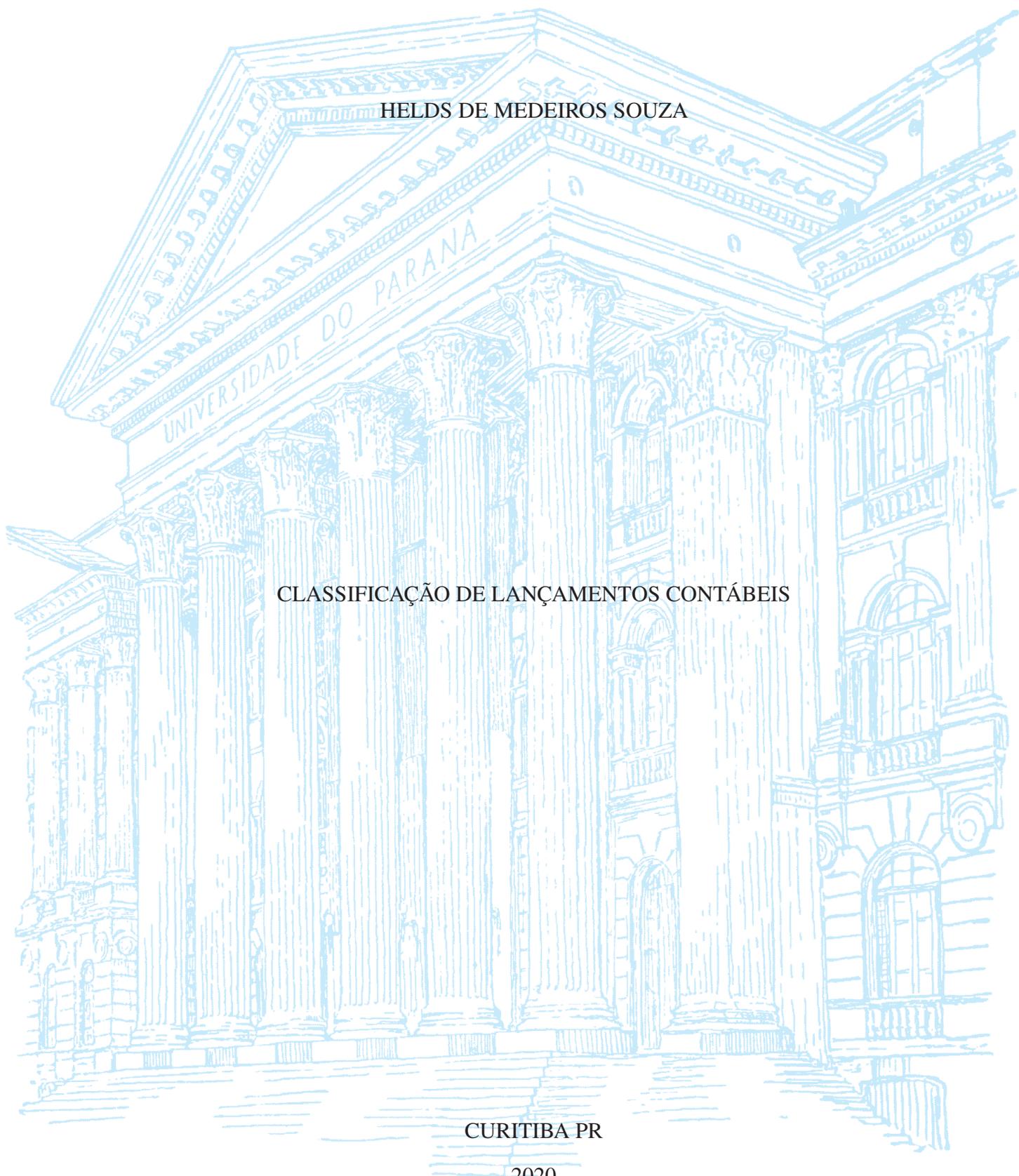
UNIVERSIDADE FEDERAL DO PARANÁ

HELDS DE MEDEIROS SOUZA

CLASSIFICAÇÃO DE LANÇAMENTOS CONTÁBEIS

CURITIBA PR

2020



HELDS DE MEDEIROS SOUZA

CLASSIFICAÇÃO DE LANÇAMENTOS CONTÁBEIS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Renato Carmo.

Coorientador: André Guedes.

CURITIBA PR

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

S729c Souza, Helder de Medeiros
Classificação de lançamentos contábeis / Helder de Medeiros Souza –
Curitiba, 2020.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de
Ciências Exatas, Programa de Pós-Graduação em Informática.

Orientador: Renato José da Silva Carmo.

Coorientador: André Luiz Pires Guedes.

1. Contabilidade Processamento de dados. I. Universidade Federal
do Paraná. II. Carmo, Renato José da Silva. III. Guedes, André Luiz
Pires. IV. Título.

CDD: 657.0285

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **HELDS DE MEDEIROS SOUZA** intitulada: **Classificação de lançamentos contábeis**, sob orientação do Prof. Dr. RENATO JOSÉ DA SILVA CARMO, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 27 de Março de 2020.

Assinatura Eletrônica

29/05/2020 10:45:42.0

RENATO JOSÉ DA SILVA CARMO
Presidente da Banca Examinadora

Assinatura Eletrônica

03/06/2020 08:45:49.0

MARCOS SFAIR SUNYE
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

28/05/2020 13:48:33.0

ANDRÉ LUIZ PIRES GUEDES
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

29/05/2020 14:10:54.0

CLAUDIO MARCELO EDWARDS BARROS
Avaliador Externo (DEPARTAMENTO DE CONTABILIDADE - UFPR)

RESUMO

Este trabalho propõe um método de classificação para lançamentos contábeis que auxilia o usuário da contabilidade na identificação de padrões para entregar análises mais tempestivas e com qualidade. Através das classes, o usuário pode analisar as transações ocorridas em um determinado período de tempo, sem a necessidade de conhecimento prévio das transações ou relatórios extracontábeis. O método utiliza as contas contábeis do lançamento para realizar a classificação.

Foi realizada uma aplicação em dados de uma empresa multinacional onde foi possível identificar facilmente alterações de padrões ocorridas na contabilidade da companhia. Através da frequência das classes o usuário da contabilidade pode identificar transações que não são realizadas de forma regular na empresa. Para classes com bastante frequência é possível realizar uma análise conjunta com outras variáveis e identificar mudanças de comportamentos. Pela simplicidade do método qualquer usuário de posse do livro diário e plano de contas pode aplicá-lo.

Palavras-chave: Contabilidade, Classificação, Fraude, Erro

ABSTRACT

This work propose a classification method for accounting postings to support the accounting users to identify patterns and provide quality and timely analysis. Using the classes the user can analyse the transaction occurred in a certain period of time, without need of previous knowledge of the transactions or non-accounting reports.

The method was applied in a multinational company where was identified patterns changes and non-recurring postings. The relative frequency of the classes could be used to identify transactions not common in the company. Classes with enough items could be analysed together with others variables to identify change of behavior in the company. Due to the simplicity of the method any users with the accounting general ledger and chart of accounts can run the analysis.

Keywords: Accounting, Classification, Fraud, Error

LISTA DE FIGURAS

4.1	Hierarquia do exemplo de plano de contas	16
4.2	Hierarquia do plano de contas do exemplo	22
5.1	Painel de análise das classificações	26
5.2	Evolução mensal do total de lançamentos	27
5.3	Painel da classe do Nível 0 do plano de contas	27
5.4	Número de classes geradas por período	28
5.5	Apresentação da classe 206 por mês	29
5.6	Apresentação da classe 206 por origem	30
5.7	Apresentação da classe 206 por dia do mês.	30
5.8	Apresentação da classe 206 por trimestre	31

LISTA DE TABELAS

4.1	Exemplo de plano de contas	17
4.2	Exemplo de lançamentos no livro diário	18
4.3	Mapeamento das contas do lançamento 3, do nível 5 para o 3	19
4.4	Lançamentos com as contas substituídas por contas do Nível 3	20
4.5	Lançamento 3 com agrupamento das contas a créditos iguais	20
4.6	Relevância das contas a débito do lançamento 3	20
4.7	Relevância das contas do lançamento 3	21
4.8	Partidas do lançamento 3 após exclusão da conta 3.3.1	21
4.9	Estrutura do identificador de lançamento.	21
4.10	Identificador de lançamento dos lançamentos	22
4.11	Classe com taxa de ocorrência.	22
5.1	Número de contas por nível do plano de contas.	23
5.2	Quantidade de partidas agrupadas por nível do plano de contas	24
5.3	Número de classes por nível do plano de contas	24
5.4	Distribuição do número de partidas por lançamento	25

SUMÁRIO

1	INTRODUÇÃO	8
2	DEFINIÇÕES BÁSICAS DE CONTABILIDADE	10
3	REVISÃO DE LITERATURA	11
3.1	MÉTODOS ANALÍTICOS PARA ANÁLISE DE ERROS E FRAUDES	11
3.2	MÉTODOS COMPUTACIONAIS PARA IDENTIFICAÇÃO DE OBSERVAÇÕES DISCREPANTES	12
3.3	CRÍTICA SOBRE OS MÉTODOS PESQUISADOS	13
4	MÉTODO PROPOSTO	15
4.1	ENTRADAS E PARÂMETROS DA CLASSIFICAÇÃO	15
4.1.1	Plano de contas	17
4.1.2	Livro diário	18
4.1.3	Parâmetros	18
4.2	PREPARAÇÃO	19
4.2.1	Substituição do nível das contas das partidas	19
4.2.2	Consolidação de partidas do mesmo tipo e com mesma conta	19
4.2.3	Cálculo da relevância das contas	20
4.2.4	Ordenação das contas	21
4.2.5	Exclusão de contas não relevantes	21
4.3	IDENTIFICADOR DE LANÇAMENTO	21
4.4	CLASSIFICAÇÃO DOS LANÇAMENTOS CONTÁBEIS	22
4.4.1	Descrição das classes	22
4.4.2	Distribuição de frequência de classes	22
5	APLICAÇÃO EM DADOS REAIS	23
5.1	CONJUNTO DE DADOS	23
5.2	DEFINIÇÃO DO PLANO DE CONTAS	23
5.3	PROCESSAMENTO DA CLASSIFICAÇÃO	24
5.3.1	Efeito do plano de contas na classificação	25
5.3.2	Efeito do corte de partidas não relevantes na classificação	25
5.3.3	Painel de análise das classificações	25
5.4	ANÁLISE DAS CLASSES GERADAS E RESULTADOS OBTIDOS	26
6	CONCLUSÃO	32
6.1	TRABALHOS FUTUROS	32
	REFERÊNCIAS	34

1 INTRODUÇÃO

A contabilidade tem por objetivo registrar e analisar o patrimônio das entidades e a suas alterações. Para Marion, a contabilidade é um instrumento que fornece o máximo de informações úteis para a tomada de decisões, dentro e fora da empresa (Marion, 2009). O autor ainda reforça que os usuários da contabilidade são “todas as pessoas ou entidades interessadas em conhecer a situação da empresa para tomada de decisões, como administradores, gerentes, governo, bancos, fornecedores, etc”.

As demonstrações contábeis podem conter informações incorretas, levando os usuários a tomarem decisões diferentes daquelas que tomariam caso estas estivessem corretas. Os erros na contabilidade podem ser causados por negligência, imprudência, imperícia ou ainda por fraude.

Lançamentos contábeis fictícios são frequentemente realizados para encobrir atos ilícitos. Para detectar estes tipos de lançamentos, a implementação de canais de denúncias, a execução de testes documentais ou de consistência entre relatórios são indicados. A prevenção de fraude também inclui a implementação de programas de ética e governança, além de controles que impeçam ou desestimulem que fraudes sejam executadas.

As disciplinas de auditoria e controles internos foram as que mais desenvolveram metodologias para análise das informações contábeis, a fim de identificar possíveis fontes de erros ou fraudes. Softwares para a auditoria como o ACL (ACL, 2018) e IDEA (Analytics, 2018) fornecem diversas técnicas de análise de dados como classificação, estratificação, análise de dígitos (Benford's Law) (Nigrini, 2012), teste de lançamentos duplicados, análise de “gap”, procura de termos, entre outras análises (Spann, 2013). Estas técnicas não consideram a estrutura do lançamento. Muitos dos métodos são aplicados em relatórios extracontábeis, como o relatório de títulos a pagar, por exemplo.

Técnicas de identificação de observações discrepantes podem servir para identificar lançamentos não usuais. Hawkins define que uma observação discrepante é uma observação que desvia tanto das demais que levanta a suspeita de que foi gerada por um mecanismo diferente (Hawkins, 1980). Aggarwal lista uma série de técnicas para a identificação de observações discrepantes (Aggarwal, 2013). Estas técnicas, entretanto, exigem que pressupostos, como a independência das observações por exemplo, sejam atendidos para que funcionem. Como um lançamento contábil é representado por várias partidas no livro diário, para executar os métodos seria necessário realizar uma transformação da entrada. Este tipo de procedimento no livro diário pode trazer problemas citados por Aggarwal, como matrizes esparsas, onde a maioria dos números são preenchidos por zeros, prejudicando a qualidade dos resultados.

Alguns outros métodos também podem ser utilizados para a identificação de observações discrepantes. Análise de agrupamentos identifica subgrupos ou grupos em um conjunto de dados (James et al., 2013). Neste método são buscadas partições entre as observações de forma que, dentro de cada grupo formado, as observações sejam parecidas entre si, e que sejam diferentes de outros grupos. As partições são criadas levando em consideração o número de agrupamentos que se deseja obter e a distância entre uma observação e as demais.

Os métodos de classificação são utilizados para prever variáveis qualitativas (Aggarwal, 2013). Para isto são calculados coeficientes de regressão, a partir de observações previamente classificadas. Caso uma classe não esteja definida ou não haja nenhuma observação desta classe nas observações classificadas previamente, os métodos de classificação podem apresentar problemas e classificar observações em uma classe na qual elas não pertencem naturalmente.

O método proposto por este trabalho é uma função que classifica os lançamentos contábeis a partir das contas utilizadas em seu registro. Esta função recebe as partidas do lançamento, o plano de contas da empresa, e os parâmetros “Nível do Plano de Contas”, no qual pretende-se fazer a classificação e o “Ponto de Corte” para a eliminação de contas não relevantes. O resultado desta função é a classe do lançamento contábil, realizada a partir da construção de um identificador da transação. O resultado do método não substitui nem automatiza o trabalho do usuário da contabilidade, ele serve para ampliar a capacidade de análise e julgamento profissional. Através da análise do tamanho da classe, o usuário da contabilidade pode identificar transações que não são realizadas com muita frequência. Para cada nova classe criada o usuário da contabilidade deve dar um nome que represente as transações desta classe, facilitando a identificação do tipo de lançamentos contidos na classe. Devido a simplicidade da execução do método apresentado, ele pode ser aplicado por qualquer usuário da contabilidade que tenham acesso ao livro diário e plano de contas da companhia.

O texto deste trabalho está organizado da seguinte maneira. O Capítulo 2 apresenta conceitos básicos da contabilidade que são utilizados pelo método. O Capítulo 3 mostra alguns métodos de auditoria para a análise de erros e fraudes na contabilidade, além de métodos estatísticos e computacionais para a identificação de valores discrepantes em conjuntos de dados. O Capítulo 4 apresenta o método para a classificação de lançamentos contábeis e a identificação de classes não usuais. Neste capítulo também são mostrados os formatos de alguns dos relatórios utilizados neste trabalho. No Capítulo 5 são realizadas análises das classes geradas a partir da aplicação do método em dados contábeis de uma empresa multinacional, no período de janeiro de 2018 até abril de 2019. Por fim, no Capítulo 6 são apresentadas conclusões obtidas através do estudo realizado.

2 DEFINIÇÕES BÁSICAS DE CONTABILIDADE

Este capítulo apresenta alguns conceitos contábeis básicos, necessários para o entendimento do método de classificação de lançamentos contábeis, proposto neste trabalho.

Um lançamento contábil registra um evento econômico que ocorreu na entidade. Este registro é normalmente realizado pelo método das partidas dobradas, onde para cada débito existe um crédito correspondente. O crédito é a origem do recurso representado pelo lançamento, já o débito, onde este recurso foi aplicado. As partidas contábeis representam as diversas origens e aplicações dos recursos em um lançamento que afetam os itens patrimoniais. O livro diário é um relatório que contém um conjunto de lançamentos contábeis, organizados por dia. Neste relatório cada partida é uma linha, de modo que um lançamento contábil contém duas ou mais linhas.

O balanço patrimonial apresenta os ativos e passivos da entidade em uma determinada data. Ativos são bens e direitos, enquanto os passivos são obrigações contraídas pela entidade. A demonstração do resultado do exercício (DRE) apresenta as receitas e despesas da entidade, apuradas em um determinado período de tempo. Ativos e despesas têm natureza devedora, enquanto passivos e receitas têm natureza credora. Os itens de natureza devedora têm o seu saldo aumentado quando recebem uma partida a débito, e o seu saldo diminuído quando recebem uma partida a crédito. O contrário ocorre para itens de natureza credora.

As contas contábeis são codificações utilizadas para representar os itens patrimoniais, ou seja, bens, direitos, obrigações, despesas e receitas. Contas que recebem o saldo de partidas contábeis são chamadas analíticas. As contas sintéticas são representações mais genéricas dos itens patrimoniais. Seus saldos são calculados a partir da somatória das contas analíticas que tem a mesma natureza que a conta sintética. Este agrupamento pode ser feito de modo recursivo e forma vários níveis de agrupamentos. Um plano de contas é o conjunto de todas as contas disponíveis para serem utilizadas em uma entidade. Nele também é definido como os agrupamentos das contas acontece.

O método das partidas dobradas garante que em cada lançamento, o total de créditos é igual ao total de débitos. Em vários sistemas informatizados de contabilidade o crédito é tratado com sinal negativo, fazendo com que a soma de débitos e créditos seja zero em todo lançamento. No balanço patrimonial a mesma propriedade faz com que a soma das contas do ativo seja igual a soma das contas do passivo.

As contas são organizadas por ordem de liquidez, tanto no balanço patrimonial quanto no plano de contas. Liquidez é o quão rápido um bem ou direito é convertido em dinheiro, por exemplo, quão rápido uma dívida pode ser exigida pelos credores.

Algumas técnicas para análise de balanço são:

Análise horizontal: Verifica-se a variação das contas de um período para outro. Esta análise pode ser realizada em termos absolutos ou percentuais.

Análise vertical Verifica-se a representatividade da conta em relação ao grupo a qual pertence.

Análise de quocientes Calcula-se razões entre saldos de balanço e compara com padrões de mercado ou com o histórico destes índices da própria empresa.

3 REVISÃO DE LITERATURA

Neste capítulo é apresentada a pesquisa bibliográfica a respeito dos métodos existentes para análise de erros e fraude na contabilidade. Também são apresentados métodos de agrupamento e classificação utilizados na computação e estatística para a identificação de observações discrepantes (*outliers*). A pesquisa bibliográfica é dividida em duas partes. Na Seção 3.1 são apresentados métodos aplicados pela disciplina de auditoria para a identificação de erros ou fraudes. Na Seção 3.2 são apresentados alguns métodos para a identificação de padrões e anomalias aplicados pela ciência de dados.

3.1 MÉTODOS ANALÍTICOS PARA ANÁLISE DE ERROS E FRAUDES

Para Bourke a utilização de *Computer Aided Audit Tools* (CAAT's), sigla em inglês que significa técnicas de auditoria assistida por computador, está se tornando popular entre os profissionais de auditoria e examinadores de fraude (Bourke, 2010). Para o autor estas técnicas são utilizadas para simplificar ou automatizar o processo de análise de dados. O autor destaca que a sigla também se tornou sinônimo de incorporar análise de dados no processo de auditoria.

Bourke ainda lista algumas técnicas que podem ser realizadas utilizando CAATs, são elas:

- Filtro/Visualização baseada em critério
- Histórico (*Aging*)
- Expressões/Equações
- União/Relações
- Análise de lacunas (*gaps*)
- Análise de tendências
- Análise estatística
- Análise de regressão
- Duplicidades
- Simulação paralela
- Ordenação/Indexação
- Benford's *Law*
- Sumarização
- Concordância (*Matching*)
- Estratificação
- Combinação de várias análises

Rosplock descreve alguns processos que podem ser utilizados para detectar fraudes (Rosplock, 2001).

1. Realizar uma análise subjetiva do histórico de operações da companhia. Identificar mudanças de hábitos de pagamentos, aplicações financeiras, exposição a empréstimos de curto/longo prazo.

2. Analisar a condição financeira através da análise horizontal e vertical do balanço patrimonial e da demonstração do resultado do exercício. Comparar a condição financeira e dos quocientes com os padrões do mercado.

3. Realizar uma análise de tendência dos índices financeiros. Segundo o autor isto pode ajudar a identificar padrões inconsistentes nos quocientes e estatísticas.

4. A detecção de inconsistências nas tendências requer mais análises para determinar os fatores que impactaram nas mudanças das condições dos quocientes e estatísticas financeiras.

No livro *Fraud Analytics* Spann introduz sete softwares que implementam CAATS para a análise de fraudes (Spann, 2013). Os softwares apresentados no livro são:

1. ACL Analytics 10
2. CaseWare IDEA
3. Raytheon's VisuaLinks
4. IBM's i2 Analyst's Notebook
5. Centrifuge Visual Network Analytics.
6. SAS Analytics
7. Actionable Intelligence Technologies Financial Investigative Software

Thiprungsri descreve um método de agrupamento que utiliza a descrição dos lançamentos contábeis de contas transitórias de um banco internacional como forma de identificar padrões não usuais nesses lançamentos (Thiprungsri, 2012).

3.2 MÉTODOS COMPUTACIONAIS PARA IDENTIFICAÇÃO DE OBSERVAÇÕES DISCREPANTES

Aggarwal afirma que em muitas aplicações os dados são gerados por um ou mais processos geradores e que quando o processo gerador se comporta de maneira não usual isto resulta na criação de uma observação discrepante (Aggarwal, 2013).

Erros podem fazer com que o sistema de contabilidade se comporte de maneira não usual e assim gerar observações discrepantes. Técnicas que identificam observações discrepantes nos lançamentos contábeis então podem auxiliar os profissionais de contabilidade para identificar tais erros e entregar análises com melhor qualidade.

No livro *Outlier Analysis*, Aggarwal apresenta alguns tipos de modelos para a identificação de observações discrepantes. Abaixo segue um resumo dos principais métodos apresentados pelo autor.

Na análise de valores extremos, são considerados como observações discrepantes valores muito pequenos ou muito grandes. A distribuição dos valores é presumida e as probabilidades são calculadas em suas caudas. Esta análise é bem simples para dados uni-variados, porém em espaços multidimensionais pode não apresentar um bom desempenho. Este tipo de análise, entretanto, acaba sendo muito utilizada para avaliar os valores extremos dos escores de anormalidades, calculados por outros métodos de detecção (Aggarwal, 2013).

Nos modelos probabilísticos, a distribuição de probabilidade é suposta, e os parâmetros do modelo são calculados a partir dos valores das observações. Uma vantagem destes métodos é que podem ser aplicados para dados contínuos, ordinais ou categóricos. Como a distribuição é suposta, ela pode não se ajustar perfeitamente aos dados. Para dimensões altas, o modelo pode ser ajustado pelas observações discrepantes e não ser capaz de identificá-las. Muitos modelos paramétricos são difíceis de interpretar e podem não explicar o processo de geração de dados anormais (Aggarwal, 2013).

Os modelos lineares pertencem a uma dimensão menor do que a dos dados. Por exemplo, dados em um espaço de dimensão dois, podem ser modelados através de uma reta. Normalmente

o método dos mínimos quadrados é utilizado para determinar um subespaço dimensional menor. Neste caso são calculados os resíduos dos dados em relação ao modelo, e utilizado a análise de valores extremos para determinar as observações discrepantes (Aggarwal, 2013).

Os modelos baseados em distâncias calculam similaridade ou dissimilaridade entre as observações. A ideia é considerar os dados isolados dos demais como observações discrepantes. Estes métodos podem ser divididos entre Análise de Agrupamentos (*Cluster Analysis*), Métodos Baseados em Densidade (*Density-based methods*) e o Método dos k-Vizinhos Mais Próximos (*k-nearest neighbors*). Aggarwal afirma que estes métodos tipicamente podem ser computacionalmente caros (Aggarwal, 2013).

A ideia do Modelo baseado na Teoria da Informação é que as observações discrepantes aumentam o tamanho mínimo do “código” requerido para descrever o conjunto de dados. Este código pode ser medido como entropia, complexidade de Kolmogorov, variância etc. São consideradas discrepantes observações que ao serem retiradas do conjunto de dados reduzem de forma significativa a medida utilizada para descrever os dados (Aggarwal, 2013).

A identificação de observações discrepantes em dimensões altas é desafiadora. A esparsidade aumenta de forma considerável nestas dimensões e os dados se tornam equidistantes. Do ponto de vista da densidade quase todas as regiões são igualmente esparsas, considerando todas as dimensões. Nestes casos, as observações discrepantes são normalmente identificadas em projeções sobre subespaços de dimensões menores. Neste tipo de análise são utilizadas tanto técnicas de agrupamentos quanto de regressão. Estes métodos podem ser computacionalmente difíceis de se implementar (Aggarwal, 2013).

Souza apresentou uma metodologia para a detecção de lançamentos não usuais na contabilidade (Souza, 2017). O autor utiliza um método de agrupamentos hierárquicos e detecta quais lançamentos “demoram” mais para agrupar-se com os demais.

Antes de calcular a similaridade entre os lançamentos contábeis, Souza realiza diversos procedimentos para reduzir a dimensão e complexidade dos dados (Souza, 2017). Dois deles, considerados relevantes para esta pesquisa, é a eliminação do efeito das variações do montante no cálculo da medida de distância e a identificação de lançamentos idênticos através da concatenação de texto.

O primeiro procedimento divide o valor da partida pelo total de débito do documento. O outro ordena as partidas e faz uma concatenação de texto das contas utilizadas, juntamente com o percentual das partidas. Ao final do segundo procedimento, as partidas são ordenadas de acordo com os textos concatenados, e os lançamentos que têm os mesmos textos são agrupados.

Apesar do autor ter concluído que o método ajuda a identificar lançamentos não usuais na contabilidade, os critérios para a escolha da medida de similaridade e o ponto de corte dos agrupamentos não são muito claros.

3.3 CRÍTICA SOBRE OS MÉTODOS PESQUISADOS

Os métodos citados na Seção 3.1 demandam um conhecimento prévio do contexto operacional da empresa. Alguns dos métodos também exigem que sejam requisitadas informações extracontábeis para a conclusão sobre inconsistências.

Alguns métodos de detecção de fraude utilizam procedimentos de aprendizado de máquina para identificar atributos de transações sabidamente fraudulentas para assim determinar novos casos de fraude. No entanto, identificar casos de fraude não é um trabalho fácil, e não são todas as empresas que têm esse histórico para fornecer.

Os métodos computacionais e estatísticos para a identificação de dados discrepantes são complexos de se implementar e demandam um alto conhecimento estatístico para a interpretação

de seus resultados. Estes métodos também exigem que algumas premissas sejam atendidas para sua efetividade. Além disso, dados não independentes, esparsos e com multicolinearidade são difíceis de modelar. Os resultados obtidos na aplicação dos métodos apresentados para tais observações podem não ser confiáveis.

Em métodos de agrupamento baseados em distância é necessário definir com antecedência o número de grupos que se pretende obter. Isto pode fazer com que os grupos naturais de transações não sejam descobertos. O valor das distâncias fica distorcido quando há esparsidade no conjunto de dados.

Em métodos de classificação é necessário ter conhecimento das classes antes da execução do método. Caso apareçam dados que pertençam a classes não incluídas nas observações previamente classificadas, essas classes podem não ser descobertas.

O trabalho realizado por Souza tem os mesmos problemas apresentados por Aggarwal no que tange a métodos de agrupamento baseados em distâncias. Porém projeções realizadas no conjunto de dados e o método de comparação de lançamentos idênticos podem ser utilizados neste trabalho.

4 MÉTODO PROPOSTO

O método proposto por este trabalho é uma função que classifica os lançamentos contábeis a partir das contas utilizadas em seu registro. Esta função recebe as partidas do lançamento, o plano de contas da empresa, e os parâmetros Nível do Plano de Contas, no qual pretende-se fazer a classificação e o Ponto de Corte para a eliminação de contas não relevantes. O resultado desta função é a classificação do lançamento contábil, realizada a partir da construção de um identificador da transação.

É importante ressaltar que a classificação não depende de outros lançamentos do livro diário para ser realizada. Para cada lançamento é contruído um identificador a partir das contas utilizadas para o seu registro e a classificação dos lançamentos se dá pela simples comparação por igualdade dos identificadores, não sendo necessário o cálculo de similaridade, conforme os métodos tradicionais de classificação apresentados na Seção 3.2. O método também não depende de informações extracontábil, que são utilizadas por alguns procedimentos de auditoria apresentados na Seção 3.1.

Os registros contábeis das transações são obtidos a partir do livro diário e as informações referentes às contas utilizadas são obtidas do plano de contas. A Seção 4.1 apresenta a estrutura do livro diário e do plano de contas. Estes relatórios podem ser facilmente obtidos pelo usuário da contabilidade da empresa. Além disto, em sistemas informatizados de contabilidade, eles seguem um certo padrão.

Os procedimentos necessários para padronizar os lançamentos contábeis, a fim de obter classes que sejam significativas para a análise, são explicados na Seção 4.2. Nesta padronização, o usuário da contabilidade deve definir alguns parâmetros, tais como o nível do plano de contas a ser utilizado e o limite que define se uma conta é ou não relevante para a transação.

A Seção 4.3 explica como o identificador de lançamento é construído. Ele foi criado para representar a essência da estrutura do lançamento contábil e permitir que os lançamentos possam ser comparados através de comparação por igualdade. O procedimento utiliza a concatenação de texto das contas das partidas a crédito, seguidas pelas contas das partidas a débito, ordenadas por relevância.

Na Seção 4.4 é explicado como deve ser feito as descrições das classes geradas pelo método de classificação. Também é apresentado como a distribuição de frequência das classes é utilizada para determinar se a classe é ou não usual nas transações da entidade.

4.1 ENTRADAS E PARÂMETROS DA CLASSIFICAÇÃO

Nesta seção são apresentados o plano de contas, o livro diário e os parâmetros necessários para a execução do método de classificação proposto por este trabalho.

Foi criado um plano de contas com 57 contas, das quais 22 são analíticas, que é utilizado nos lançamentos da Tabela 4.2. O plano de contas pode ser visto na Tabela 4.1. Este plano contém cinco níveis, de 1 a 5, e foi adicionado o nível 0 que agrupa as contas “Ativo”, “Passivo” e “Resultado”.

Este plano de contas é uma estrutura hierárquica. Em cada nível, as contas são agrupadas de acordo com sua natureza, e não por sua similaridade. Por exemplo, a conta analítica “1.1.1.2.2” é a conta “Banco ag432 cc83495”, mas ela também é uma conta “Banco” bem como “Caixa e equivalente de caixa”. A Figura 4.1 mostra as mesmas contas em uma estrutura de árvore, com o nível 0 na raiz e as contas analíticas, de nível 5 nas folhas.

Tabela 4.1: Exemplo de plano de contas

Número	Descrição	Nível	Tipo	Conta Superior
0	Partida dobrada	0	C	
1	Ativo	1	S	0
1.1	Circulante	2	S	1
1.1.1	Caixa e equivalente de caixa	3	S	1.1
1.1.1.1	Caixa	4	S	1.1.1
1.1.1.1.1	Caixa	5	A	1.1.1.1
1.1.1.2	Bancos	4	S	1.1.1
1.1.1.2.1	Banco ag123 cc78643	5	A	1.1.1.2
1.1.1.2.2	Banco ag432 cc83495	5	A	1.1.1.2
1.1.1.3	Aplicações	4	S	1.1.1
1.1.1.3.1	Aplicação financeira	5	A	1.1.1.3
1.1.2	Clientes	3	S	1.1
1.1.2.1	Clientes	4	S	1.1.2
1.1.2.1.1	Clientes a receber	5	A	1.1.2.1
1.1.3	Estoque	3	S	1.1
1.1.3.1	Estoque para revenda	4	S	1.1.3
1.1.3.1.1	Estoque de mercadorias para revenda	5	A	1.1.3.1
1.2	Não Circulante	2	S	1
1.2.1	Imobilizado	3	S	1.2
1.2.1.1	Custo	4	S	1.2.1
1.2.1.1.1	Imóveis	5	A	1.2.1.1
1.2.1.1.2	Máquinas e equipamentos	5	A	1.2.1.1
1.2.1.1.3	Veículos	5	A	1.2.1.1
1.2.1.2	Depreciação acumulada	4	S	1.2.1
1.2.1.2.1	(-) Depreciação Acumulada Imóveis	5	A	1.2.1.2
1.2.1.2.2	(-) Depreciação Acumulada Máq e Equip	5	A	1.2.1.2
1.2.1.2.3	(-) Depreciação Acumulada Veículos	5	A	1.2.1.2
2	Passivo	1	S	0
2.1	Circulante	2	S	2
2.1.1	Contas a pagar	3	S	2.1
2.1.1.1	Fornecedores	4	S	2.1.1
2.1.1.1.1	Fornecedores	5	A	2.1.1.1
2.1.1.2	Obrigações sociais e trabalhistas	4	S	2.1.1
2.1.1.2.1	Salários	5	A	2.1.1.2
3	Resultado	1	S	0
3.1	EBITDA	2	S	3
3.1.1	Receita	3	S	3.1
3.1.1.1	Receita	4	S	3.1.1
3.1.1.1.1	Receita na venda de mercadorias	5	A	3.1.1.1
3.1.1.1.2	Receita na venda de imobilizado	5	A	3.1.1.1
3.1.1.2	Devoluções	4	S	3.1.1
3.1.1.2.1	Devolução de vendas	5	A	3.1.1.2
3.1.2	Custo	3	S	3.1
3.1.2.1	Custo	4	S	3.1.2
3.1.2.1.1	Custo na venda de imobilizado	5	A	3.1.2.1
3.1.2.1.2	Custo na venda de mercadorias	5	A	3.1.2.1
3.1.3	Outras despesas	3	S	3.1
3.1.3.1	Outras despesas	4	S	3.1.3
3.1.3.1.1	Despesa com locação de equipamentos	5	A	3.1.3.1
3.2	Depreciação/Amortização	2	S	3
3.2.1	Depreciação	3	S	3.2
3.2.1.1	Depreciação	4	S	3.2.1
3.2.1.1.1	Despesa com depreciação	5	A	3.2.1.1
3.3	Resultado financeiro	2	S	3
3.3.1	Despesas financeiras	3	S	3.3
3.3.1.1	Despesas financeiras	4	S	3.3.1
3.3.1.1.1	Despesa com juros	5	A	3.3.1.1

Para explicar o método foi criado um exemplo de livro diário contendo quatro lançamentos, conforme mostrados na Tabela 4.2. Os lançamentos utilizam as contas do plano de contas apresentado na Tabela 4.1. Os três primeiros são pagamentos de fornecedores e o quarto refere-se a pagamento de salários.

4.1.1 Plano de contas

Para a classificação, o plano de contas deve conter todas as contas em seu nível mais analítico e ir agrupando-as, até que no nível mais alto, restem apenas Ativo, Passivo e Resultado. O agrupamento das contas no plano deve ser monotônico, ou seja, uma vez que duas contas são agrupadas elas não se separam em níveis superiores do plano de contas.

Somente o código da conta, o nível e sua conta agrupadora são utilizados pelo método de classificação. Pode-se optar por incluir outras variáveis informativas, como descrição da conta, tipo da conta, entre outras, sem prejuízo para a execução do método.

Tabela 4.2: Exemplo de lançamentos no livro diário

Lcto.	Linha	Conta	Descrição da Conta	Descrição do Lcto.	Tipo	Montante
1	1	1.1.1.2.1	Banco ag123 cc78643	Pagamento de fornecedor	C	-3.450
1	2	2.1.1.1.1	Fornecedores	Pagamento de fornecedor	D	3.450
2	1	1.1.1.2.1	Banco ag123 cc78643	Pagamento de fornecedor	C	-23.000
2	2	2.1.1.1.1	Fornecedores	Pagamento de fornecedor	D	23.000
3	1	1.1.1.2.2	Banco ag432 cc83495	Pagamento de fornecedor	C	-12.000
3	2	1.1.1.1.1	Caixa	Pagamento de fornecedor	C	-3.500
3	3	2.1.1.1.1	Fornecedores	Pagamento de fornecedor	D	15.000
3	4	3.3.1.1.1	Despesa com juros	Pagamento de fornecedor	D	500
4	1	1.1.1.2.1	Banco ag123 cc78643	Pagamento de salários	C	-35.000
4	2	2.1.1.2.1	Salários	Pagamento de salários	D	35.000

4.1.2 Livro diário

Além das informações básicas da transação, o livro diário pode conter uma série de outras informações gerenciais. O método de classificação proposto neste trabalho utiliza somente as informações listadas a seguir:

Número do documento: Número único que identifica um lançamento contábil.

Tipo da partida: Identifica se a partida é do tipo devedora ou credora. Pode ser identificado também pelo sinal do valor da partida.

Número da conta: Código da conta contábil.

Valor: Informa o valor da partida. A soma de todas as partidas a débito é igual a soma das partidas a crédito para cada documento.

Entretanto, para compreender e descrever as classes criadas o usuário da contabilidade pode incluir outras informações como data, descrição, usuário, etc.

Nos lançamentos de exemplo da Tabela 4.2, foram considerados o número do documento, o número da partida, a conta, a descrição da conta, o tipo da partida, a descrição da transação e o valor das partidas.

4.1.3 Parâmetros

Em alguns casos, o usuário da contabilidade pode querer considerar o identificador de lançamento igual para lançamentos com uma pequena variação. Ele pode querer considerar por exemplo um pagamento de fornecedor com juros igual a um pagamento de fornecedor sem juros, caso os juros não sejam significativos para a operação. Para possibilitar esta flexibilização, um dos passos do método de classificação elimina contas não relevantes para o lançamento. O ponto de corte para eliminar partidas não relevantes deve ser informado pelo usuário da contabilidade antes da aplicação do método. Todas as contas são consideradas, caso um ponto de corte não seja definido.

Dependendo do nível de abstração da análise pretendido pelo usuário da contabilidade ele pode escolher diferentes níveis do plano de contas para executar a classificação dos lançamentos. Como as contas são agrupadas em níveis mais sintéticos, menos grupos são criados na classificação. No nível 0 por exemplo, onde todas as contas são agrupadas em uma só, há somente uma classe.

4.2 PREPARAÇÃO

As contas utilizadas para a criação do identificador de lançamento poderão ser de qualquer nível do plano de contas. Durante a execução do procedimento de classificação, o usuário da contabilidade deve informar o nível em que deseja realizar as análises.

Como o identificador de lançamento é um texto criado a partir das contas informadas nas partidas ou nas suas correspondentes agrupadoras no plano de contas, a ordem em que elas são consideradas é importante. Contas não relevantes para descrever a transação também podem ser utilizadas, porém não representam o fato econômico ocorrido. No método de classificação, contas não relevantes ou não ordenadas farão com que mais classes sejam criadas. Esta seção mostra os procedimentos para evitar esse tipo de situação.

Nesta seção também é apresentado o procedimento de substituição das contas analíticas, que são utilizadas no livro diário, por suas respectivas contas sintéticas em outros níveis. Também é apresentado como é realizada o agrupamento de partidas do mesmo tipo e com as mesmas contas contábeis. A Seção 4.2.3 apresenta o cálculo da relevância das contas dentro de lançamento contábil e a Seção 4.2.4 como as contas serão ordenadas utilizando essa métrica. Por fim, a Seção 4.2.5 apresenta como contas não relevantes são excluídas do método de classificação.

4.2.1 Substituição do nível das contas das partidas

O nível de conta utilizado na classificação é um parâmetro que deve ser fornecido pelo usuário da contabilidade. Para exemplificar a classificação utilizamos contas do Nível 3. Utilizando a estrutura do plano de contas, conforme ilustrado na Figura 4.1, é obtida a respectiva conta no Nível 3, a partir da conta do Nível 5 informada no livro diário. A Tabela 4.3 mostra as respectivas contas no Nível 3 para o Lançamento 3.

Tabela 4.3: Mapeamento das contas do lançamento 3, do nível 5 para o 3

Conta 5	Descrição da Conta 5		Conta 3	Descrição da Conta 3
1.1.1.2.2	Banco ag432 cc83495	→	1.1.1	Caixa e Equivalentes de Caixa
1.1.1.1.1	Caixa	→	1.1.1	Caixa e Equivalentes de Caixa
2.1.1.1.1	Fornecedores	→	2.1.1	Contas a Pagar
3.3.1.1.1	Despesa com juros	→	3.3.1	Despesas Financeiras

A Tabela 4.4 mostra os lançamentos com as contas substituídas pelas contas de Nível 3. É possível verificar que após a substituição de contas, os lançamentos 1, 2 e 4 se tornaram bastante similares, tendo como diferenças somente o valor de suas partidas.

4.2.2 Consolidação de partidas do mesmo tipo e com mesma conta

O método de classificação proposto por este trabalho utiliza a relevância da conta e o tipo da partida (credora ou devedora). Sendo assim, quando existem partidas do mesmo tipo e com a mesma conta, elas são agrupadas. Desta forma a relevância da conta considera os valores de todas as partidas na qual ela foi utilizada.

Devido ao procedimento da Seção 4.2.1 no Lançamento 3, as partidas 1 e 2, que são do tipo credora, têm suas contas substituídas pela “1.1.1”. Para que a conta não fique duplicada no identificador, o valor de todas as partidas do mesmo tipo são somados. Depois deste procedimento o Lançamento 3 fica conforme a Tabela 4.5.

Tabela 4.4: Lançamentos com as contas substituídas por contas do Nível 3

Lcto.	Linha	Conta	Descrição da Conta	Descrição do Lcto.	Tipo	Montante
1	1	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de fornecedor	C	-3.450
1	2	2.1.1	Contas a Pagar	Pagamento de fornecedor	D	3.450
2	1	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de fornecedor	C	-23.000
2	2	2.1.1	Contas a Pagar	Pagamento de fornecedor	D	23.000
3	1	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de fornecedor	C	-12.000
3	2	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de fornecedor	C	-3.500
3	3	2.1.1	Contas a Pagar	Pagamento de fornecedor	D	15.000
3	4	3.3.1	Despesas Financeiras	Pagamento de fornecedor	D	500
4	1	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de salários	C	-35.000
4	2	2.1.1	Contas a Pagar	Pagamento de salários	D	35.000

Tabela 4.5: Lançamento 3 com agrupamento das contas a créditos iguais

Lcto.	Conta	Descrição da Conta	Descrição do Lcto.	Tipo	Montante
3	1.1.1	Caixa e Equivalentes de Caixa	Pagamento de fornecedor	C	-15.500
3	2.1.1	Contas a Pagar	Pagamento de fornecedor	D	15.000
3	3.3.1	Despesas Financeiras	Pagamento de fornecedor	D	500

4.2.3 Cálculo da relevância das contas

Seja V_C o conjunto dos valores das contas a crédito de um determinado lançamento. A relevância da i -ésima conta a crédito R_{Ci} é dada pela Equação 4.1, que divide o valor movimentado na conta i , V_{Ci} , pelo valor da conta com o maior valor absoluto a crédito no lançamento, $MaxV_C$. Utilizamos o módulo dos valores a créditos pois eles estão representados com sinal negativo no livro diário utilizado.

$$R_{Ci} = \frac{V_{Ci}}{MaxV_C} \quad (4.1)$$

Da mesma forma o valor da relevância da j -ésima conta a débito, R_{Dj} , é dado pela Equação 4.2.

$$R_{Dj} = \frac{V_{Dj}}{MaxV_D} \quad (4.2)$$

Para a conta a crédito do Lançamento 3 o $MaxV_C$ é igual a 15.500 e a relevância R_C foi calculada utilizando a Equação 4.1. Como há apenas uma partida a crédito neste lançamento a relevância é igual a 100%.

Para as contas a débito do mesmo lançamento $MaxV_D = 15.000$, R_D foi calculado utilizando a Equação 4.2 a relevância de cada partidas é mostrada na Tabela 4.6.

Tabela 4.6: Relevância das contas a débito do lançamento 3

Lcto.	Conta	Descrição da Conta	V_D	R_D
3	2.1.1	Contas a Pagar	15.000	100%
3	3.3.1	Despesas financeiras	500	3%

4.2.4 Ordenação das contas

Como o identificador de lançamento é uma sequência ordenada de contas, a ordem que as contas são listadas no livro diário altera o resultado da classificação. Para ser comparáveis, todos os lançamentos devem seguir os mesmos critérios de ordenação de suas contas. Utilizamos a relevância das contas no lançamento como critério de ordenação. Para contas com mesma relevância, consideramos os códigos das contas em ordem alfanumérica como critério de ordenação.

Tabela 4.7: Relevância das contas do lançamento 3

Lcto.	Conta	Descrição da Conta	Tipo	Montante	R
3	1.1.1	Caixa e Equivalente de Caixa	C	-15.500	100
3	2.1.1	Contas a Pagar	D	15.000	100
3	3.3.1	Despesas Financeiras	D	500	3

4.2.5 Exclusão de contas não relevantes

Neste exemplo são consideradas somente contas com relevância maior ou igual a 33%. A conta “3.3.1” no Lançamento 3^o tem relevância igual a 3, conforme pode ser visto na Tabela 4.7. Sendo assim, excluímos esta conta para fins de classificação do lançamento 3. Na Tabela 4.8 a conta não relevante foi eliminada.

Tabela 4.8: Partidas do lançamento 3 após exclusão da conta 3.3.1

Lcto.	Conta	Descrição da Conta	Tipo	Montante	R
3	1.1.1	Caixa e Equivalente de Caixa	C	-15.500	100
3	2.1.1	Contas a Pagar	D	15.000	100

4.3 IDENTIFICADOR DE LANÇAMENTO

O identificador de lançamento é uma sequência ordenada das contas que representam as origens e as aplicações dos recursos movimentados pela entidade. Sua ordenação deve ocorrer conforme descrito na Seção 4.2.4. Ele é construído após os passos explicados na Seção 4.2.

No exemplo utilizado, a sequência de contas foi concatenada em um texto. Esta representação, no entanto, pode ser implementada de diversas outras formas. Na primeira parte foi adicionada a sequência de contas a crédito, separadas por ponto e vírgula. Depois disto, foi concatenado o símbolo de arroba seguido da sequência de contas a débito, separadas por ponto e vírgula.

A Tabela 4.9 mostra as três partes do exemplo de identificador de lançamento aplicado ao Lançamento 3 utilizando o Nível 5 do plano de contas.

Tabela 4.9: Estrutura do identificador de lançamento

Código das contas a crédito	1.1.1.2.2;1.1.1.1.1
Separador de Tipo de partida	@
Código das contas a débito	2.1.1.1.1;3.3.1.1.1

4.4 CLASSIFICAÇÃO DOS LANÇAMENTOS CONTÁBEIS

Para definir a classe do lançamento, basta verificar o identificador de lançamento. Ele é que define qual a classe do lançamento contábil. Como o identificador de lançamento pode ser diferente, dependendo dos parâmetros utilizados para sua construção, um mesmo lançamento pode pertencer a classes diferentes.

Tabela 4.10: Identificador de lançamento dos lançamentos

Lcto.	Identificador
1	1.1.1@2.1.1
2	1.1.1@2.1.1
3	1.1.1@2.1.1
4	1.1.1@2.1.1

Conforme pode ser verificado na Tabela 4.10, todos os lançamentos utilizados no exemplo, considerando o Nível 3 do Plano de Contas, possuem o mesmo identificador de classe. É importante ressaltar que este resultado depende do nível da conta utilizado e do ponto de corte para as exclusão de partidas não relevantes, definidos na Seção 4.2.

4.4.1 Descrição das classes

Para cada nova classe criada, o usuário da contabilidade deve realizar a descrição da classe. Para isto ele deve utilizar as contas que compõem a classe e como elas se relacionam, conforme Figura 4.2, além do histórico de cada lançamento incluído na classe.

Figura 4.2: Hierarquia do plano de contas do exemplo



Para a classe “1.1.1@2.1.1” foi dada a descrição “Pagamento de fornecedores”.

4.4.2 Distribuição de frequência de classes

Larson define a frequência f como a contagem do número de entradas em uma determinada classe. A frequência relativa é definida como a porcentagem de dados que está em uma determinada classe, e pode ser calculada dividindo a frequência f pelo tamanho da população analisada n , conforme equação 4.3.

$$f_R = \frac{f}{n} \quad (4.3)$$

Tabela 4.11: Classe com taxa de ocorrência

Identificador	Descrição	Ocorrência
1.1.1@2.1.1	Pagamento de fornecedores	100%

Utilizamos a frequência relativa para identificar classes com poucas observações. Desta forma é possível apontar para o usuário da contabilidade as transações que não são usuais.

5 APLICAÇÃO EM DADOS REAIS

Neste capítulo apresentamos uma aplicação do método proposto utilizando um livro diário de uma empresa multinacional, do período de janeiro de 2018 até abril de 2019. Além de classificar os lançamentos contábeis, realizamos análises utilizando outros atributos dos lançamentos contábeis.

Na Seção 5.1 apresentamos estatísticas descritivas do conjunto de dados. Na Seção 5.2 definimos o parâmetro do plano de contas. Na Seção 5.3 são realizadas os agrupamentos das contas, de acordo com o parâmetro nível da conta e construído o identificador do lançamento. Também é abordado sobre o efeito do parâmetro nível da conta e ponto de corte na classificação. Na Seção 5.4 são realizadas análises das classes e discussões sobre os resultados obtidos nesta aplicação.

5.1 CONJUNTO DE DADOS

O livro diário utilizado contém um total de 186.334 lançamentos contábeis. O período contemplado pelo relatório vai de janeiro de 2018 ao início de abril de 2019. O número de linhas desse conjunto de dados é de 464.664, pois cada lançamento contém duas ou mais partidas.

Um total de 418 contas analíticas foram movimentadas nesse período. Dessas contas 116 são do ativo, 172 de custos e despesas, 87 de passivo e 43 de receitas.

Estes lançamentos juntos movimentaram um total de R\$34.259.339.169, 23. Este valor representa o montante movimentado entre contas e não o valor gerado pela companhia no período.

5.2 DEFINIÇÃO DO PLANO DE CONTAS

O plano de contas é uma entrada no método de classificação proposto por este trabalho. A classificação ocorre no nível desejado. Para um melhor controle dos resultados obtidos na aplicação do método, criamos um plano de contas genérico.

O plano de contas criado tem estrutura hierárquica com cinco níveis sintéticos, de 4 até 0. As contas analíticas, utilizadas no livro diário, são de Nível 5 e são mapeadas para as contas do Nível 4.

São utilizadas 418 contas analíticas para o registro das transações no conjunto de dados obtidos. No Nível 4 estas contas são agrupadas em 74 contas sintéticas. No Nível 3 o número de contas reduz para 45. Na Tabela 5.1 vemos o número de contas para cada nível.

Tabela 5.1: Número de contas por nível do plano de contas

Nível.	Nr de Contas
0	1
1	2
2	9
3	45
4	74
5	418

As contas criadas no Nível 3 retratam os grupos de balanço e resultado que normalmente são apresentados em demonstrações financeiras. Neste nível é possível abstrair alguns detalhes e avaliar a situação dos itens de balanço e resultado de maneira mais holística.

Para comparar o efeito do plano de contas na classificação analisamos os identificadores de lançamento utilizando os níveis 3 e 4.

5.3 PROCESSAMENTO DA CLASSIFICAÇÃO

O protótipo implementado utiliza a linguagem SQL em algumas etapas.

Efetuada o agrupamento das contas no Nível 5, o número total de partidas reduziu para 399.615. No Nível 0 todas as contas se agrupam em uma só no plano de contas de forma que o número de partidas agrupadas é o dobro do número de lançamentos. Na Tabela 5.2 verificamos o número de partidas agrupadas resultantes da aplicação do agrupamento, considerando as contas em cada um dos níveis do plano de conta.

Tabela 5.2: Quantidade de partidas agrupadas por nível do plano de contas

Nível	Nro. Partidas
5	399.615
4	381.887
3	381.577
2	377.524
1	377.385

Para executar o classificador organizamos as informações pela seguinte ordem:

- a) Número do documento;
- b) Identificador de débito e crédito;
- c) Relevância;
- d) Conta.

Para o classificador é utilizado somente o número do documento, Identificador de débito e crédito e a Conta. Partidas com relevância inferior a 33% foram desconsideradas.

Para armazenar o resultado da classificação criamos duas tabelas. Uma para as classes criadas e outra para relacionar os documentos com o classes para um determinado nível do plano de contas. Para cada novo identificador de classe criamos um índice na tabela de classes. Em seguida adicionamos as classes e os respectivos documentos que foram alocados para esta classe em um dado nível do plano de contas.

No total foram geradas 2.382 classes, distribuídas conforme a Tabela 5.3 pelos níveis do plano de contas.

Tabela 5.3: Número de classes por nível do plano de contas

Nível.	Nr de Classes
0	1
1	38
2	103
3	370
4	501
5	1.372

5.3.1 Efeito do plano de contas na classificação

Após o agrupamento de contas, os lançamentos com mais de duas partidas no Nível 5 do plano de contas representam somente 3,4% do número total de lançamentos. Aqueles com mais de cinco partidas representam 0,9%. Quando as partidas são agrupadas no Nível 3 do plano de contas o percentual de lançamentos com mais de cinco partidas cai para 0,1%.

Tabela 5.4: Distribuição do número de partidas por lançamento

Partidas	Qtd	Qtd %
2	181.421	97,4
3	2.854	1,5
4	490	0,3
5	1.477	0,8
>=6	92	0

No conjunto de dados com o qual realizamos a análise, executamos o método de classificação considerando dois planos de contas no Nível 3. O número de classes da primeira execução foi de 462, já no plano de contas que foi refeito o número de classes reduziu 20%, ficando em 370.

5.3.2 Efeito do corte de partidas não relevantes na classificação

Para fins de classificação foram eliminadas todas as partidas com relevância menor ou igual a 33%. Este procedimento fez com que a representatividade dos lançamentos com mais de duas partidas caísse para 1,8% com as contas no Nível 5 e para 0,5% para o agrupamento nas contas de Nível 3.

O agrupamento de contas tem uma participação importante na determinação de partidas que serão excluídas para fins da classificação. O número de partidas com menos de 33% de relevância no Nível 5 chega a 25.580. Quando é considerado este mesmo ponto de corte para partidas agrupadas no Nível 3, o número de partidas excluídas cai para 8.243. Quanto menos partidas nos lançamentos, menor será o número de classes criadas. o usuário da contabilidade pode utilizar este parâmetro para controlar o grau de detalhes que gostaria de analisar.

5.3.3 Painel de análise das classificações

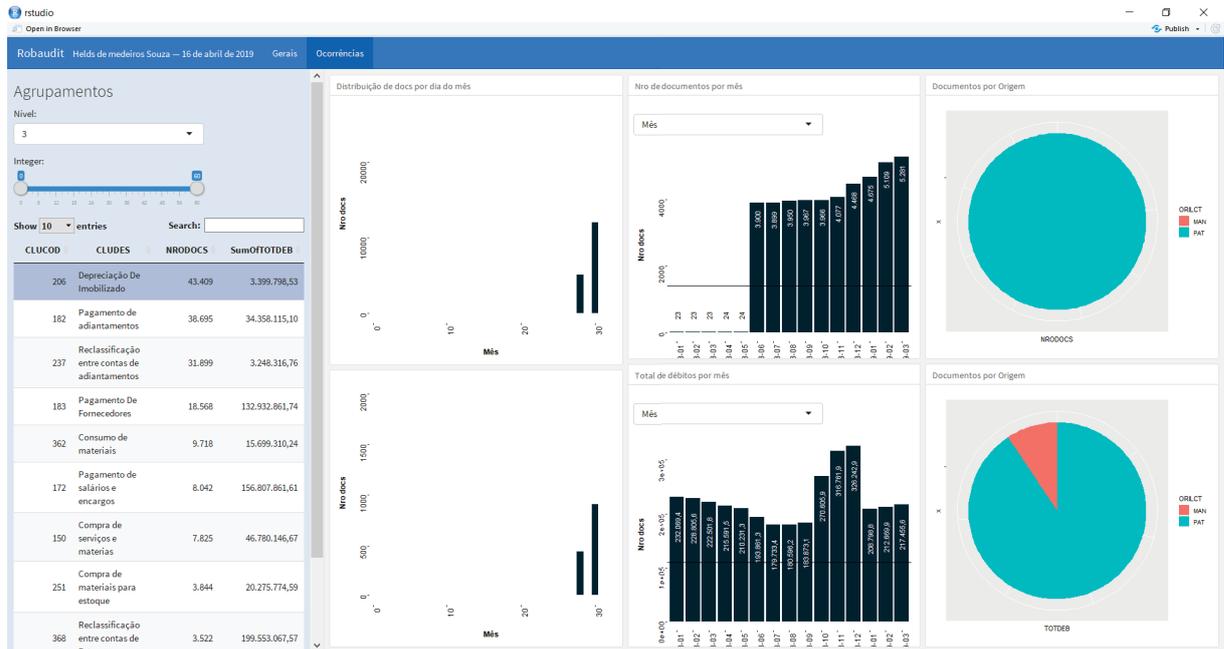
Para efetuar a análise da classificação realizada criamos um painel com informações relevantes a respeito de cada classe, conforme pode ser visto na Figura 5.1. Neste painel são listadas todas as classes, de acordo com o nível do plano de contas selecionado. Uma vez definido o nível do plano de contas é mostrado na faixa lateral uma tabela com os identificadores de classe, suas descrições, o total de documentos agrupados e o total de montante movimentado.

Quando um identificador de classe é selecionado são exibidos no painel seis gráficos. Os gráficos da primeira linha consideram a quantidade de lançamentos nas classes, já os da segunda linha consideram o total de débitos movimentados para apresentar algumas estatísticas.

Na primeira coluna é mostrada a quantidade de lançamentos e o total de débitos movimentados distribuídos pelos dias do mês. Estes gráficos servem para mostrar os dias do mês em que os lançamentos ocorrem com mais frequência ou para verificar em quais dias do mês ocorrem as movimentações mais relevantes.

Na segunda coluna temos a evolução da quantidade e do total de débitos movimentados por mês ou por trimestre, dependendo da escolha do analista. Estes gráficos mostram a evolução e alterações de comportamentos da companhia.

Figura 5.1: Painel de análise das classificações



A terceira coluna do painel mostra a distribuição da quantidade de documentos e do total de débitos movimentados, abertos por origem dos lançamentos. Estes gráficos mostram qual módulo do sistema ERP gerou os lançamentos contábeis. O objetivo destes gráficos é mostrar se há a preponderância de uma única origem dos lançamentos.

5.4 ANÁLISE DAS CLASSES GERADAS E RESULTADOS OBTIDOS

Na análise das classes geradas verificamos as contas que foram afetadas pelos lançamentos através do identificador de lançamento. Esta verificação nos permite avaliar a conta da origem dos recursos e onde eles foram aplicados. Em seguida verificamos o histórico individual de cada lançamento da classe. Na análise individual dos lançamentos classificados é possível verificar o texto de cada lançamento em particular, o que facilita a descrição da classe como um todo. Como é realizada uma análise individual da classe, para classes com poucas observações já é possível que o usuário da contabilidade identifique registros não usuais.

Na Figura 5.2(a) verificamos que há uma tendência de crescimento no número de documentos por mês. Se considerarmos a evolução do total de débitos, apresentado na Figura 5.2(b), não identificamos nenhuma tendência, já que há uma discrepância muito grande do total de débitos movimentados em dezembro de 2018 para os demais meses. Esta discrepância se deve ao zeramento do resultado do exercício que acontece anualmente.

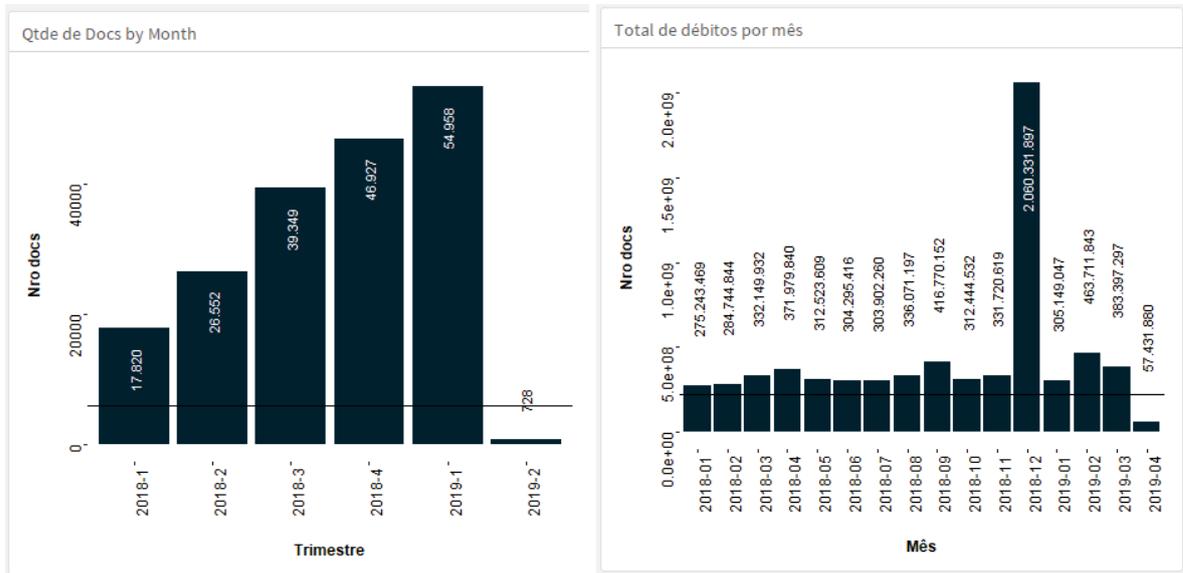
No Nível 0 do plano de contas existe apenas uma classe, já que neste nível todas as contas são agrupadas em uma só, e todos os lançamentos têm origem e aplicação nesta mesma conta. Na Figura 5.3 é possível verificar um pico discreto do número de documentos no final do mês. Já a maior movimentação em relação ao total de débitos ocorre no final do mês.

No gráfico de evolução da quantidade de lançamentos por mês é possível verificar que há uma tendência de crescimento do número de lançamentos contábeis. Na evolução do total de débito movimentado por mês é possível verificar um pico ocorrido no período 2018-12. Este pico foi devido ao lançamento do zeramento anual do resultado que ocorre neste período.

Figura 5.2: Evolução mensal do total de lançamentos

(a) Quantidade de lançamentos

(b) Soma do total de débitos



Os lançamentos com origem no módulo de contas a pagar (PAG) representam mais de 50% do número de documentos registrados, seguido dos lançamentos de patrimônio (PAT) que representam pouco menos que 25%. Da perspectiva do total de débitos, os lançamentos com origem manual (MAN) representam mais de 60% do total de débitos movimentados.

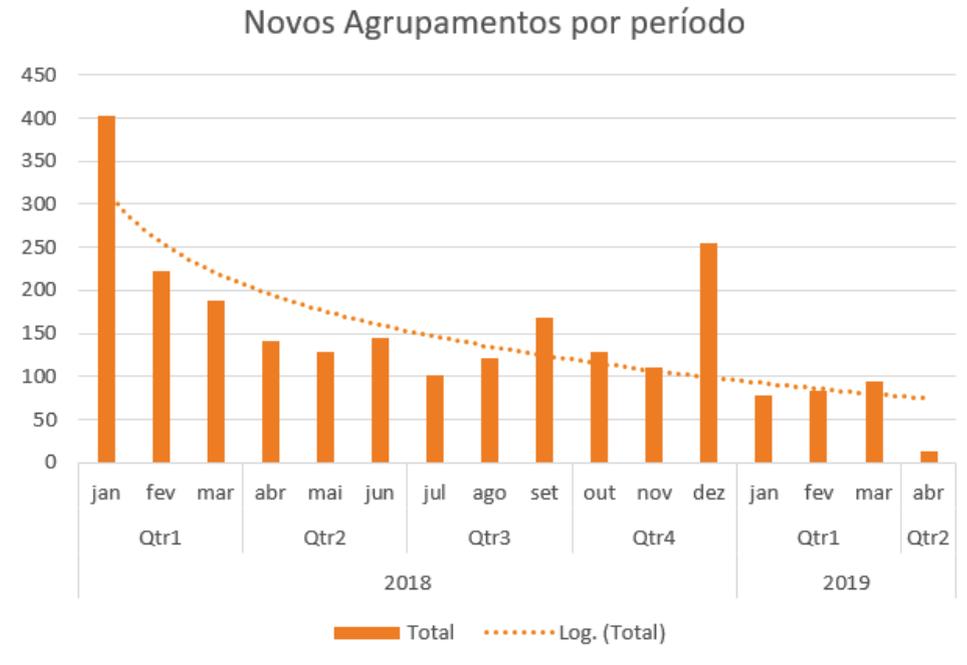
Figura 5.3: Painel da classe do Nível 0 do plano de contas



Na Figura 5.4 verificamos uma redução no número de novas classes criadas por período. Em dezembro de 2018 há um volume significativo de novas classes, devido a transações que acontecem anualmente. Entretanto, também foram feitos diversos ajustes contábeis para assegurar a correta apresentação do balanço da empresa, o que resultou em novas classes. Percebemos também que a geração de novas classes é maior no último mês de cada trimestre. Este comportamento só não foi observado no primeiro trimestre de 2018 onde os lançamentos mais

usuais ainda estavam sendo classificados. Isto acontece porque nos trimestre existem revisões mais criteriosas dos lançamentos e mais ajustes são lançados.

Figura 5.4: Número de classes geradas por período



Através da análise do painel identificamos mudanças de comportamento nas atividades da empresa. Analisando a classe do Nível 3 de número 206, por exemplo, visualizamos uma mudança de padrão do mês de maio de 2019 para o mês subsequente. No gráfico da Figura 5.5(a) o número sai de 24 em 2018-05 para 3.900 em 2018-06. Questionada a administração informou que no mês da mudança de comportamento a pessoa responsável pela atividade foi substituída e a pessoa que assumiu a atividade passou a contabilizar a despesa com depreciação por bem, e não de forma agrupada como era realizado até então.

Já o gráfico da Figura 5.5(b) não indica um aumento de 2018-05 para 2018-06. O que mostra que de fato o valor dos 24 lançamentos simplesmente foi diluído nos 2.900. Já nos períodos 2018-10, 2018-11 e 2018-12 verificamos um aumento nos totais de débitos movimentados. Analisando os registros identificamos que no período 2018-10 foi realizado um ajuste manual para acelerar a depreciação de um bem específico, que não foi possível corrigir no sistema a tempo do fechamento mensal. Nos períodos 2018-11 e 2018-12 este mesmo evento foi contabilizado, porém através do sistema.

Este único lançamento manual também chamou atenção no gráfico da Figura 5.6 que mostra a quantidade e o total de débitos por origem dos lançamentos. No gráfico da Figura 5.6(a) é possível constatar que quase 100% dos lançamentos desta classe têm origem no módulo de patrimônio (PAT), entretanto a categoria de lançamentos manuais representa cerca de um oitavo do total de débitos movimentados, conforme verificamos na Figura 5.6(b).

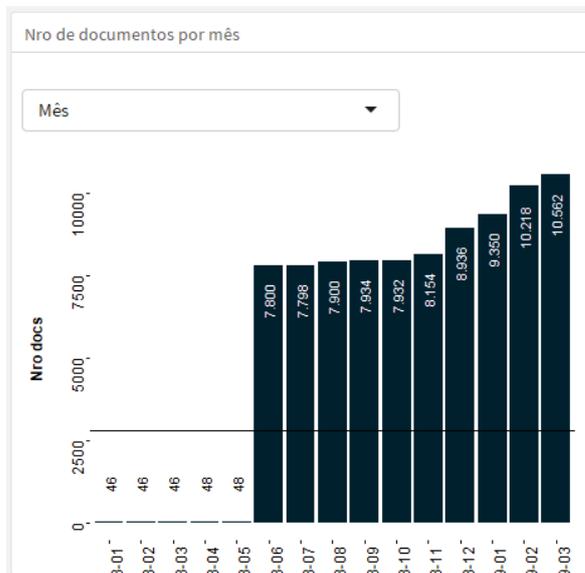
Na Figura 5.7 não identificamos nenhuma anomalia. É possível confirmar o que esperávamos para este grupo de lançamentos, que é a concentração dos lançamentos nos últimos dias do mês, já que o cálculo de depreciação é realizado e contabilizado no final de cada mês.

No painel também analisamos a evolução do número de documentos e do total de débitos movimentados agrupado por trimestre.

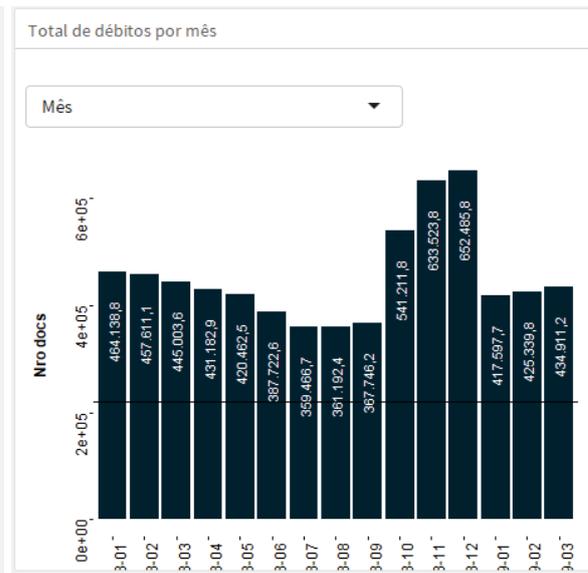
Nesta análise da classificação fizemos diversos achados nas transações contábeis. Abaixo fizemos uma lista dos itens mais relevantes.

Figura 5.5: Apresentação da classe 206 por mês

(a) Quantidade de lançamentos



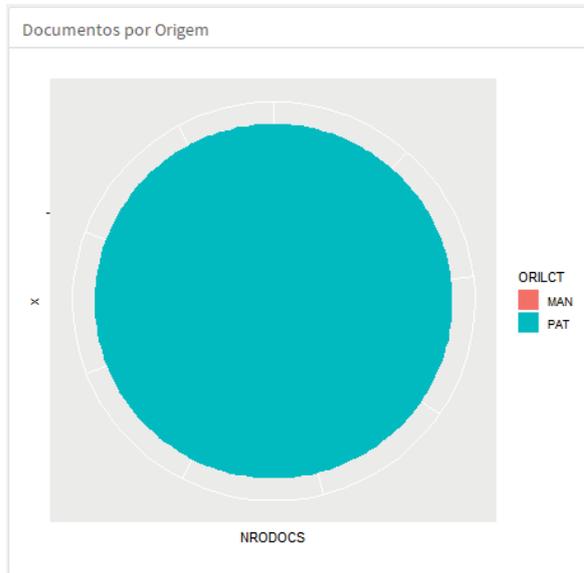
(b) Total de débitos



- Tendência de crescimento do número de lançamentos por mês;
- Relevância dos lançamentos de zeramento de exercício em dezembro de 2018;
- Maior concentração de lançamentos no final do mês;
- Lançamentos da origem PAG (Pagamentos) tem maior número de ocorrências, porém lançamentos de origem MAN (Manuais) registram os maiores valores;
- O número de novas classes criadas reduz ao longo do tempo;
- O número de novas classes é maior nos meses de final de trimestre;
- Mudança de comportamento da contabilização de depreciação na Classe 206;
- Pico de valor na Classe 206 nos meses de outubro à dezembro de 2018;

Figura 5.6: Apresentação da classe 206 por origem

(a) Quantidade de lançamentos



(b) Total de débitos

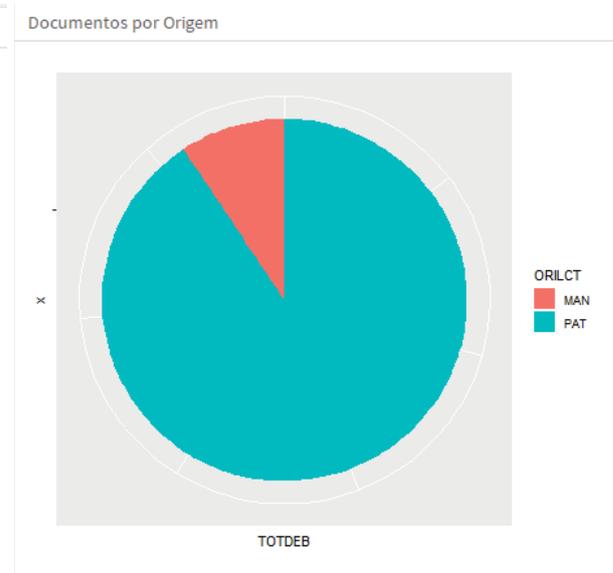
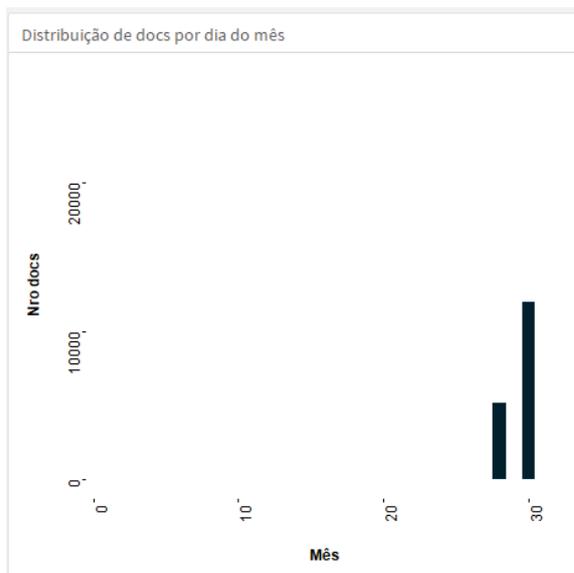


Figura 5.7: Apresentação da classe 206 por dia do mês

(a) Quantidade de lançamentos



(b) Total de débitos

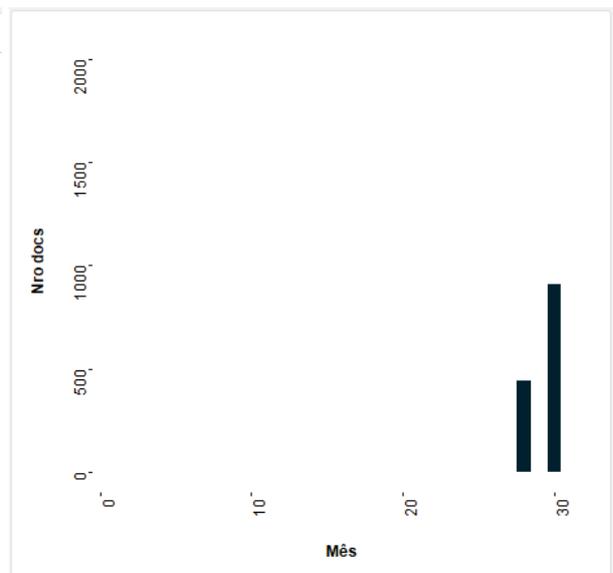
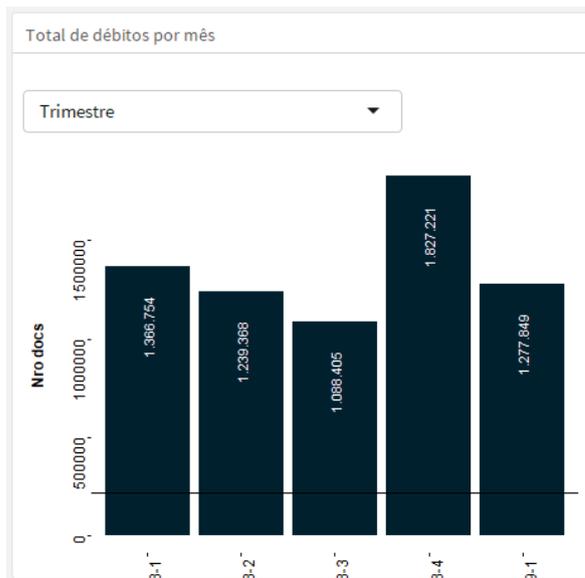
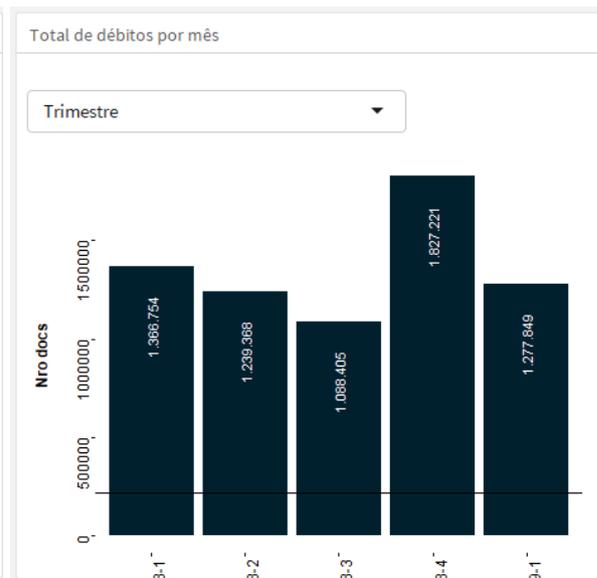


Figura 5.8: Apresentação da classe 206 por trimestre

(a) Quantidade de lançamentos



(b) Total de débitos



6 CONCLUSÃO

A administração e outros profissionais que atestam os lançamentos contábeis precisam de métodos para assegurar que as demonstrações financeiras estão livres de erros materiais. O método proposto neste trabalho é uma ferramenta que contribui para este fim. Ele também desenvolve a pesquisa de métodos quantitativos aplicados à contabilidade e aproxima a Ciência da Computação da Ciência Contábil.

Na aplicação prática verificamos que as classes criadas através do método proposto por este trabalho facilitam a análise dos lançamentos contábeis. Quando o número de documentos em uma classe é suficientemente grande, é possível extrair informações que antes estavam ocultas. Para classes com poucas observações, no momento de dar nome à classe já é possível realizar uma análise crítica dos mesmos e identificar possíveis erros e fraudes.

Para as classes com apenas um lançamento, as análises de comportamento não são aplicáveis. Neste caso a própria análise realizada pelo usuário da contabilidade para dar nome à classe já é suficiente para avaliar a veracidade do lançamento.

Verificamos que quando um lançamento tem muitas partidas, algumas delas podem não ser relevantes nos níveis mais analíticos. Porém quando as contas são agrupadas em um nível superior do plano de contas, uma nova partida pode se tornar relevante, diferenciando a classe. Sendo assim, lançamentos de uma mesma classe, em um nível mais analítico do plano de contas, podem estar em classes distintas em um nível mais sintético do plano de contas.

Notamos que é difícil dar um nome que represente todos os documentos de forma razoável nos níveis 0, 1 e 2 do Plano de Contas. Os níveis 3 e 4 do Plano de Contas foram os que apresentaram um resultado mais satisfatório. Nestes níveis é possível caracterizar a natureza da operação realizada sem entrar em detalhes, trazendo vantagens para análises gerenciais.

O processo de classificação dos lançamentos contábeis foi implementado em um caso de teste na empresa que utilizamos para realizar as análises deste trabalho. Percebemos que após um período inicial, onde as classes das transações mais frequentes foram criadas, as novas classes criadas tratavam-se ou de lançamentos não usuais ou da correção de transações não usuais.

Concluimos então que é possível auxiliar o usuário da contabilidade na identificação de erros e fraudes através da classificação dos lançamentos contábeis utilizando o método proposto por este trabalho.

No planejamento estratégico as classes podem auxiliar na projeção dos efeitos das ações estratégicas da companhia em seu patrimônio. As classes podem auxiliar também na identificação de correlações entre eventos econômicos.

As classes oriundas da aplicação do método proposto, combinado com outras variáveis, pode munir o usuário da contabilidade de uma ferramenta poderosa para a identificação de erros e fraudes na contabilidade, além de ser um aliado para o entendimento das operações financeiras da companhia, para a análise estratégica e de planejamento.

A utilização deste método permite a auditoria por processo de negócio, onde as operações são vistas como um todo, e não sob o ponto de vista do grupo de contas.

6.1 TRABALHOS FUTUROS

Listamos algumas perguntas, que podem agregar valor para a análise dos lançamentos contábeis, porém não são abordadas neste trabalho. São elas:

- Como apresentar de forma gráfica e interativa as classificações geradas pelo método?

- Como implementar um sistema de alarme para transações não usuais?
- Como confrontar as classes com informações extracontábeis?
- Como utilizar as classes para entender o fluxo do caixa dentro da companhia?
- Como utilizar as classes para conciliação das transações contábeis?
- Como utilizar as classes para definir um modelo para representar o relacionamento das contas contábeis e projetar balanço?

REFERÊNCIAS

- ACL (2018). *ACL*. ACL.
- Aggarwal, C. C. (2013). *Outlier Analysis*. Springer Publishing Company, Incorporated, 2nd edition.
- Analytics, C. (2018). *IDEA*. CaseWare Analytics.
- Bourke, J. (2010). Computer assisted audit techniques or caats. Acessado em 2018-10-15.
- Hawkins, D. (1980). *Identification of Outliers*. Monographs on applied probability and statistics. Chapman and Hall.
- James, G., Witten, D., Hastie, T. e Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Larson, R. (2010). *Estatística Aplicada*. Peaseon Prentice Hall, 4 edition.
- Marion, J. C. (2009). *Contabilidade Básica*. Atlas, 10ed edition.
- Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*, volume 586. John Wiley & Sons.
- Rosplock, M. F. (2001). Advanced forensic financial analysis. *Fraud Magazine*, November/December.
- Souza, H. (2017). Detecção de transações não usuais em lançamentos contábeis através de técnicas estatísticas multivariadas.
- Spann, D. (2013). *Fraud Analytics: Strategies and Methods for Detection and Prevention*. Wiley Corporate F&A. Wiley.
- Thihrungsri, S. (2012). *Cluster Analysis for Anomaly Detection in Accounting*.