



CURSO DE SISTEMAS DE INFORMAÇÃO

IDENTIFICAÇÃO DE GASTOS MUNICIPAIS ATRAVÉS DO USO DE MINERAÇÃO DE DADOS: ESTUDO DE CASO NO PROJETO TCE/UNITINS.

MAX PEREIRA CARVALHO

Palmas - TO

2020



CURSO DE SISTEMAS DE INFORMAÇÃO

IDENTIFICAÇÃO DE GASTOS MUNICIPAIS ATRAVÉS DO USO DE MINERAÇÃO DE DADOS: ESTUDO DE CASO NO PROJETO TCE/UNITINS.

MAX PEREIRA CARVALHO

Trabalho de Conclusão de Curso apresentado ao Curso de Sistemas de Informação da Universidade Estadual do Tocantins - UNITINS, como parte dos requisitos para a obtenção do grau de Bacharel em Sistemas de Informação, sob a orientação do professor Me. Marco Antônio Firmino de Sousa.

Palmas - TO

2020



CURSO DE SISTEMAS DE INFORMAÇÃO

IDENTIFICAÇÃO DE GASTOS MUNICIPAIS ATRAVÉS DO USO DE MINERAÇÃO DE DADOS: ESTUDO DE CASO NO PROJETO TCE/UNITINS.

MAX PEREIRA CARVALHO

Trabalho de Conclusão de Curso apresentado ao Curso de Sistemas de Informação da Universidade Estadual do Tocantins - UNITINS, como parte dos requisitos para a obtenção do grau de Bacharel em Sistemas de Informação, sob a orientação do professor Me. Marco Antônio Firmino de Sousa.

**Prof. Me. Marco Antônio Firmino de
Sousa**
Orientador

Prof. Me. Douglas Chagas da Silva
Convidado 1

Prof. Me. Janio Elias Teixeira Junior
Convidado 2

Palmas - TO
2020

Este trabalho é dedicado à minha família, pelo apoio incondicional.

Agradecimentos

Primeiramente a Deus por ter me concedido saúde e força para chegar até aqui.

Aos meus familiares que sempre me incentivaram a superar as dificuldades e seguir em frente.

Aos meus amigos de jornada, por não me deixarem desistir.

E a todos professores pelo conhecimento passado através desses anos.

*“Por isso não tema, pois estou com você;
não tenha medo, pois sou o seu Deus.
Eu o fortalecerei e o ajudarei;
eu o segurarei com a minha mão direita vitoriosa.”
(Bíblia Sagrada, Isaías 41:10)*

Resumo

Com o passar do tempo a população brasileira vem se questionando cada vez mais sobre os gastos públicos. E através de leis de acesso a informações para todos, o governo disponibiliza essas informações em um grande volume de bases de dados, se tornando incompreensível para a maioria da população. Portanto, esse trabalho tem como intuito a utilização da mineração de dados para transformar esse grande volume de informações em dados compreensíveis, utilizando técnicas e ferramentas adequadas para esse processo. A partir disso, essa pesquisa utiliza uma base gerada pelo Projeto TCE/UNITINS, algoritmo de K-Means para fazer os agrupamentos dos municípios que contém características semelhantes, regressão linear para encontrar o valor esperado do grupo e normalização dos dados. A fim para detectar relações de gastos acima de padrões estabelecidos para municípios do estado do Tocantins.

Palavras-chaves: Gasto Público, Mineração de Dados, Projeto TCE/UNITINS.

Abstract

Over time, the Brazilian population has been increasingly questioning public spending. And through access to information laws for all, the government makes this information available in a large volume of databases, making it incomprehensible to the majority of the population. Therefore, this work aims to use data mining to transform this large volume of information into understandable data, using techniques and tools suitable for this process. From this, this research uses a base generated by the TCE / UNITINS Project, K-Means algorithm to make the groupings of the municipalities that contain similar characteristics, linear regression to find the expected value of the group and normalization of the data. In order to detect spending ratios above the standards established for municipalities in the state of Tocantins.

Keywords: *Public Guest, Data Mining, Project TCE/UNITINS.*

Lista de ilustrações

Figura 1 – Processo de Descoberta do Conhecimento em Base de Dados	19
Figura 2 – Pipeline Etapas de Preparação dos Dados	29
Figura 3 – Divisão dos Municípios do Tocantins em 5 Clusters	31
Figura 4 – Comparação município de Rio da Conceição	33
Figura 5 – Comparação município de Centenário	34
Figura 6 – Comparação município de Cachoeirinha	35
Figura 7 – Comparação município de Santa Terezinha	36

Lista de tabelas

Tabela 1 – Tarefas realizadas por técnicas de mineração de dados	21
Tabela 2 – Recursos/ferramentas utilizados	26
Tabela 3 – Exemplo de campos utilizados na análise de dados	27
Tabela 4 – Coluna antes do processo de padronização	30
Tabela 5 – Coluna após o processo de padronização	30
Tabela 6 – Separação dos municípios por grupos	32
Tabela 7 – Maiores lançamentos para mês de Jul/2020 - Rio da Conceição	34
Tabela 8 – Maiores lançamentos para mês de Set/2019 - Centenário	35
Tabela 9 – Maiores lançamentos para mês de Set/2020 - Cachoeirinha	36
Tabela 10 – Maiores lançamentos para mês de Dez/2019 - Santa Terezinha	37

Lista de abreviaturas e siglas

DCBD - Descoberta de Conhecimento em Bancos de Dados

DM - Data Mining

IBGE - Instituto Brasileiro de Geografia e Estatística

KDD - Knowledge Discovery in Databases

LRF - Lei de Responsabilidade Fiscal

PIB - Produto Interno Bruto

SST - Segurança no Trabalho

TCE - Tribunal de Contas do Estado

TI - Tecnologia da Informação

UNITINS - Universidade Estadual do Tocantins

Sumário

1	INTRODUÇÃO	13
1.1	Objetivos	14
1.1.1	Objetivo Geral	14
1.1.2	Objetivos Específicos	14
1.2	Justificativa	14
2	REFERENCIAL TEÓRICO	16
2.1	Data Science	16
2.2	Tipos de Padrões	17
2.2.1	Modelagem de Previsão	17
2.2.2	Análise de Associação	18
2.2.3	Agrupamentos ou <i>Clustering</i>	18
2.2.4	Deteção de Anomalias	18
2.3	Processos para Mineração	19
2.4	Bibliotecas para Mineração de Dados em Python	22
2.4.1	Pandas	22
2.4.2	Numpy	22
2.4.3	SciPy	22
2.5	Bibliotecas para Visualizações Gráficas em Python	22
2.5.1	Matplotlib	22
2.5.2	Plotly	22
2.5.3	Seaborn	23
2.6	Bibliotecas para Aprendizado de Máquina em Python	23
2.6.1	Scikit-learn	23
2.6.2	K-Mens	23
2.7	Tribunal de Contas - TCE	23
3	METODOLOGIA	26
3.1	Materiais Utilizados	26
3.2	Coleta dos Dados	26
3.3	Análise dos Dados	27
3.4	Procedimento Realizado	28
4	RESULTADOS	31
4.1	Clusterização dos Municípios	31
4.2	Comparação dos Municípios	33

4.2.1	Análise Município de Rio da Conceição	33
4.2.2	Análise Município de Centenário	34
4.2.3	Análise Município de Cachoeirinha	35
4.2.4	Análise Município de Santa Terezinha	36
5	CONCLUSÃO	38
5.1	Trabalhos Futuros	38
	REFERÊNCIAS	39

1 Introdução

Com o passar do tempo a participação e acompanhamento da população na administração pública apresentou um grande aumento. O surgimento de novas tecnologias e leis voltadas a população, como os portais transparência do governo vem sendo formas de fiscalização de gastos públicos pelo cidadão.

Uma administração transparente permite a participação do cidadão na gestão e no controle da administração pública e, para que essa expectativa se torne realidade, é essencial que ele tenha capacidade de conhecer e compreender as informações divulgadas. (FIGUEIREDO, 2017).

Mesmo com a Lei de Acesso a Informação, nº 12.527, de 18 de novembro de 2011¹, ter dado a liberdade da população possuir o acesso aos dados públicos. O grande volume de dados que é gerado e armazenado em bases de dados se torna ilegível para grande maioria da população que queira fazer esse acompanhamento da administração pública (SILVA, 2018).

E a forma de extrair essas informações de gastos públicos, na maioria das vezes, é extremamente cansativo e nada fácil. Principalmente quando o volume de informações armazenadas é muito grande, se tornando uma atividade humanamente impossível de extrair todas as informações úteis e necessárias. Assim surge a necessidade de uma ferramenta para auxiliar na extração de informações utilizando técnicas de mineração de dados.

Segundo (VILARINHO, 2018), mineração de dados é uma técnica utilizada para a obtenção de informações a partir de grandes quantidades de dados. Ela é capaz de analisar diferentes tipos de elementos e encontrar diferentes tipos de relações entre eles.

E pela visão de (KAMBER, 2006), a mineração de dados tem atraído muita atenção na indústria da informação e na sociedade como um todo nos últimos anos, devido à grande disponibilidade de grandes quantidades de dados. E (FREITAS, 2003) conclui que a Tecnologia da Informação torna-se extremamente útil ao processo decisório, pois possibilita coletar, processar e armazenar dados pela utilização de sistemas de informação.

Nesse contexto, o Projeto TCE/UNITINS concentrou, em uma base de dados, todos os dados de portais transparência dos municípios do estado do Tocantins. Como o volume de informações é muito elevado, essa base permite utilizar técnicas de mineração de dados para obter informações como: empenho, liquidação e pagamentos dos municípios do Tocantins.

¹ http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.html

analisar as interações de usuários com um ambiente de ensino à distância analisando a navegação dos usuários. Com o resultado do trabalho o autor contribuiu com uma taxonomia de acessos para associar padrões na execução de atividades propostas durante o curso e análise de melhores *templates* de navegação.

1.1 Objetivos

1.1.1 Objetivo Geral

Identificar gastos municipais com padrões elevados quando comparado a outros municípios semelhantes.

1.1.2 Objetivos Específicos

- Analisar os conceitos e técnicas de *data mining*;
- Verificar a aplicabilidade das técnicas de *data mining* para o processo de análise na base TCE/UNITINS;
- Gerar bases com municípios que contém características semelhantes para comparação.

1.2 Justificativa

De acordo com (ZANELLA, 2017), a análise das contas de entes públicos brasileiros é realizado tanto pelo controle interno quanto pelo controle externo, de acordo com a Constituição Federal de 1988. O Poder Legislativo é responsável pelo controle, junto com o Tribunal de Contas.

No entanto, (FIGUEIREDO, 2018), apresenta que o conceito de estrutura de governança de TI ainda é uma novidade para a maior parte das instituições públicas, mesmo sendo essa uma área crítica, sugere um motivo pelo qual o amadurecimento da gestão e da governança de TI ainda se mostra tão lento.

Para aproveitar plenamente o potencial da análise de dados e melhorar a eficiência das operações do setor público, o big data exige investimentos em pessoas e recursos, e escassez de mão-de-obra qualificada é um ponto relevante. Segundo (HEUDECKER, 2014).

Além da dificuldade de encontrar pessoal qualificado, há também os desafios técnicos que precisam ser enfrentados. Como desafios técnicos podem-se citar o subdesenvolvimento de ferramentas de software, a integração de múltiplas fontes e formatos de dados, armazenamento e acesso aos dados, além da falta de integração entre os sistemas e não estruturação dos sistemas legados (HEEKS, 2016).

Dessa forma os Tribunais de Contas passaram a ter um estágio de informatização mais avançado, porém ainda mantem vários dos seus processos com a mesma sistemática de antes, com fluxo manual, deixando de usar todo o poder tecnológico que tem em mãos.

Assim esta pesquisa é justificada pelo intuito de relacionar os possíveis gastos municipais que fogem do padrão em comparação com outros municípios que contém a mesma características, através da implantação de técnicas de *data mining* em uma organização governamental, em particular na base gerada pelo Projeto TCE/UNITINS.

2 Referencial Teórico

Neste momento, o grande volume de dados que são gerados por diferentes organizações apresentam um cenário onde é possível extrair duas observações: a necessidade de criação de um mecanismo eficiente para interpretação dos dados e a necessidade da transparência dessas informações.

Segundo o autor ([BITENCOURT, 2016](#)), a consonância ao atual momento da Administração Pública, voltado à gestão social, e a seus princípios constitucionais da publicidade e eficiência, a gestão transparente da informação pública é apontada como aspecto fundamental à ciência e compreensão das ações do Estado pelo cidadão e, em vista disso, à consecução da participação e do controle sociais. Todavia a regulação do acesso a essa informação é bastante recente no País.

De acordo com ([GOEBEL, 2018](#)) Todos estes dados “prendem” valiosas informações, como tendências e padrões, que podem ser usados para melhorar ou aperfeiçoar decisões de negócio. As ferramentas e técnicas empregadas para análise automática e inteligente destes dados armazenado em um data lake em que o volume de dados seja armazenado em sua forma natural, evitando pré processamento. Além disso, a plataforma deve ser transparente para que os usuários saibam exatamente como obtê-los. ([BATISTA, 2017](#)).

2.1 Data Science

Transformar dados gerados em informações úteis é sem dúvida muito importante, por meio dessa necessidade foi criado o conceito de Data Science, que, consiste na extração de conhecimento e informações para tomada de decisão através de uma grande base de dados ([TIME, 2015](#)). Um padrão pode ser caracterizado por probabilidades, e esses padrões são expressos em sua grande parte em regras, funções, fórmulas, e entre outras combinações.

De acordo com ([COSTA, 2005](#)), o interesse por este tipo de informação se deve principalmente ao fato de que as empresas e organizações estão coletando e armazenando grandes quantidades de dados como consequência da queda dos preços de meios de armazenamento e computadores e do aumento da capacidade de ambos. Ainda conclui que a popularização na utilização de armazém de dados, ou *data warehouses*, que são grandes bancos de dados criados para análise e suporte à decisão, tende a aumentar ainda mais a quantidade de informações disponível ([COSTA, 2005](#)).

A evolução da tecnologia da informação (TI) proporciona o surgimento de novos modelos que abrangem as novas tecnologias que são desenvolvidas ao longo do tempo.

Dentro do contexto da globalização e da nova era da informação, emerge a *cloud computing*, cujo modelo de arquitetura une características e vantagens de dois modelos de arquiteturas anteriores: *mainframe* e cliente-servidor (MEDEIROS, 2015).

Podemos entender *cloud computing*, também, como um serviço que disponibiliza recursos computacionais dedicados ou compartilhados entre vários usuários, sendo a alocação de recursos usualmente automatizada e adquirida de forma incremental, de acordo com as necessidades do usuário, tendo sua conectividade através da internet (VIEIRA, 2015).

Em relação a extração de padrões para geração de conhecimento, o processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases – KDD*) ganha destaque já que a visão de (TREVISAN, 2017) afirma que, este continua evoluindo, a partir da pesquisa em áreas como banco de dados, aprendizado de máquinas, reconhecimento de padrões, estatística, inteligência artificial e aquisição de conhecimento para sistemas especialistas.

2.2 Tipos de Padrões

As tarefas de mineração de dados são geralmente divididas em duas categorias: tarefas de previsão (preditivos) e tarefas descritivas (descritivos ou informativos). Isto de acordo com (ARONSON, 2009). De acordo com (COSTA, 2005) os padrões preditivos são encontrados para resolver o problema de prever o valor futuro ou desconhecido de um ou mais atributos do banco de dados a partir do valor conhecido dos demais atributos. Desta forma os padrões informativos e descritivos, tem como objetivo encontrar padrões relevantes de forma que seja interpretado pelo homem.

A previsão tem como principal objetivo prever um valor de algum atributo que seja baseado em outros que já tenha seu valor calculado, já a descritiva tem como objetivo mesclar padrões (tendências, trajetórias, grupos, correlações e anomalias) que resultam os relacionamentos dos dados. Para (CARVALHO, 2015) Existem 4 (quatro) técnicas gerais de mineração de dados que englobam todas as outras formas de apresentação e permitem uma visão mais global e apropriada ao assunto. São elas a modelagem de previsão, análise de associação, análise de agrupamentos e análise de anomalias. .

2.2.1 Modelagem de Previsão

A modelagem de previsão é a responsável por construir um modelo para a variável alvo utilizando variáveis com valores já calculados. Para que ocorra esse processo é necessário as seguintes duas etapas, são elas: classificação e regressão. A tarefa de classificação é usada para variáveis discretas, já a regressão é usada para variáveis contínuas (KUMAR, 2009). É muito importante analisar ambas etapas, para minimizar erros que podem ser

encontrados entre os valores previstos e os valores reais da variável buscada (alvo). Para melhor entendimento, um exemplo seria uma avaliação de um determinado paciente possuir uma certa doença baseado nos resultados de exames médicos.

2.2.2 Análise de Associação

A análise de associação é a responsável por descobrir padrões que podem ter um alto nível de confiança por apresentar características semelhantes entre os dados. Essa associação na maioria dos casos é feita através de regras de implicação ou é aplicada a subconjuntos de dados (KUMAR, 2009). Um exemplo para melhor entendimento, seria aplicar a análise de associação em um carrinho de compras, analisando quais produtos foram comprados juntos.

2.2.3 Agrupamentos ou *Clustering*

O agrupamento ou *clustering* é responsável por procurar e encontrar grupos que partilham semelhança, padrões e tendências. Nessa etapa não existe uma variável dependente, e não são consideradas características específicas dos atributos, dessa forma são agrupados os dados que possuem características semelhantes em segmentos ou *clusters*. De acordo com (KUMAR, 2009), o agrupamento ideal é aquele em que os itens com alto grau de similaridade se localizem no mesmo *cluster* e os itens com baixo grau de semelhança se encontrem em *clusters* distintos. Um exemplo de uso seria de agrupamentos de documentos.

2.2.4 Detecção de Anomalias

A detecção de anomalias é responsável por descobrir grupos por similaridade de valores, ou seja, comparando seus atributos, onde as características apresentadas tenha uma diferença significativa dentre os demais dados. Para (KUMAR, 2009), o objetivo de um algoritmo de detecção de anomalias é descobrir as anomalias verdadeiras e evitar a rotulação de objetos normais como anômalos. Um exemplo seria o uso é a detecção de fraudes em cartões de crédito ou padrões incomuns de doenças.

A importância de utilizar ambos no processo de mineração de dados pode acarretar variação. Mas no contexto da descoberta de conhecimento em banco de dados, os padrões descritivos tendem a ser mais utilizado quando comparado ao processo preditivos. Em contrapartida pelo ponto de vista dos pesquisadores, esse padrão é mais difícil de se avaliar significativa dentre os demais dados (KUMAR, 2009). Ainda de acordo com o autor, o valor verdadeiro não deixa claro se há uma possibilidade de sugestão ou ação para o especialista, e também não deixa claro o quanto essa sugestão ou ação seria eficaz. Isso ocorre devido que a predição normalmente é mais utilizada quando se trata de um problema mais

claro e bem específico, sendo que a busca para a resolução desse problema através da mineração seria uma resposta específica para esse problema significativa dentre os demais dados. Já no caso da descrição, são apenas um volume de dados minerados, sendo feita a decisão final do que pode ser feito para resolver esse problema pelo especialista através das informações extraídas (KUMAR, 2009).

2.3 Processos para Mineração

Os processos fundamentais para que obtenha êxito em uma mineração utilizando fonte de dados (banco de dados, relatórios, logs de acesso e transações) se iniciam a partir de uma limpeza (consistência, preenchimento de informações, remoção de ruídos e redundâncias). Então dessa forma são gerados os repositórios organizados (*Data Marts e Data Warehouses*), que já são utilizados de diversas maneiras.

O termo Descoberta de Conhecimento em Bases de Dados busca otimizar o processo para que torne os dados de baixo nível, em conhecimento de alto nível e úteis. Tornou-se estrategicamente importante pois possibilita a produção de conhecimento a partir das grandes bases de dados, sendo útil para as grandes organizações de empresas compostas por múltiplas sub-organizações (TREVISAN, 2017). Na Figura 1 logo abaixo se encontra o processo de descoberta de conhecimento a partir de uma aplicação de *data mining* em uma base de dados.

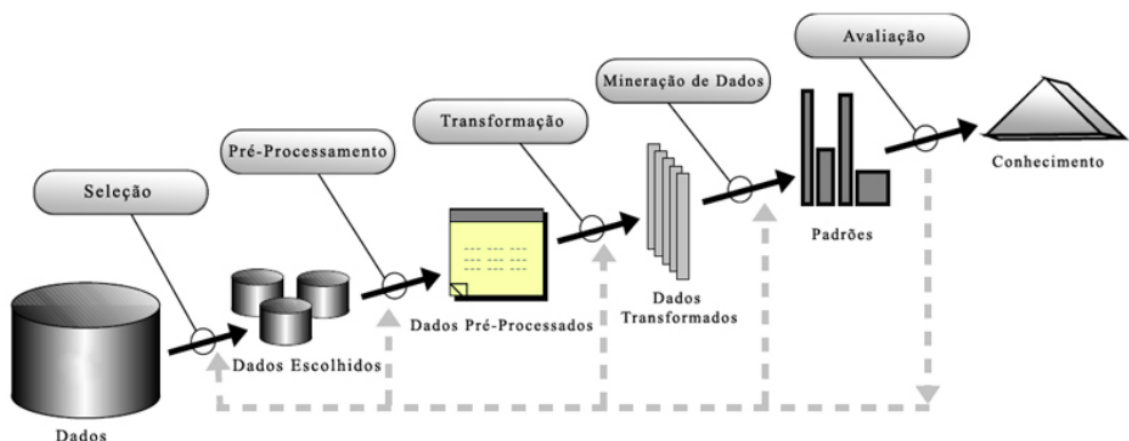


Figura 1 – Processo de Descoberta do Conhecimento em Base de Dados

Fonte: Adaptado de (FAYYAD, 1996).

Podemos observar que o processo KDD se inicia com uma escolha das fontes de dados que serão utilizados e os objetivos para a redução de dimensionamento. A partir do problema é possível aprender e desenvolver o conhecimento por meio das ferramentas capazes de extrair informações úteis em uma determinada base de dados (YAMAGUCHI, 2010).

Na segunda etapa que é o pré-processamento, ocorre a preparação dos dados, que será identificado um conjunto de dados, que fará com que ocorra uma possível realização e aplicação de técnicas de extração de conhecimentos. É nesta etapa que ocorre a busca por padronização ou modelos para o tratamento do conhecimento obtido (SOUZA, 2017).

Na terceira etapa é a mineração de dados (*Data Mining – DM*) no qual extrai padrões a partir dos dados observados. Nesta fase, são identificados os métodos e os algoritmos que irão realizar a busca pelo conhecimento que é considerado útil (SOUZA, 2017).

A mineração de dados pode ser considerada como a principal etapa de um processo de KDD, cujo o papel é incluir as tarefas de seleção, preparação e exploração das informações, e a análise e interpretação dos resultados, assimilando o conhecimento extraído do processo. Os padrões citados devem ser novos, compreensíveis e úteis, devendo trazer algum novo benefício que possa ser compreendido (COSTA, 2012).

Os resultados obtidos com a mineração de dados podem ser empregados no gerenciamento de informação, processamento de pedidos de informação, tomada de decisão, controle de processo e outras aplicações. Os dados contidos nas bases de dados são usados para aprender um determinado conceito alvo ou padrão (PASTA, 2011). Na Tabela 1 a seguir se encontra as principais tarefas de técnicas de *data mining*.

TAREFA	DESCRIÇÃO	EXEMPLO
CLASSIFICAÇÃO	Constrói um modelo de algum tipo que possa ser aplicado a dados não classificados a fim de categorizá-los em classes, tendo como objetivo relacionar o atributo meta (cujo valor será previsto) e um conjunto de atributos de previsão.	Classificar pedidos de crédito Esclarecer pedidos de seguros fraudulentos. Identificar a melhor forma de tratamento de um paciente.
ESTIMATIVA	Usada para definir um valor para alguma variável contínua desconhecida.	Estimar o número de filhos ou a renda total de uma família Estimar o valor em tempo de vida de um cliente Estimar a probabilidade de que um paciente morrerá baseando se nos resultados de diagnósticos Médicos Prever a demanda de um consumidor para um novo produto.
ASSOCIAÇÃO	Usada para determinar quais itens tendem a ser adquiridos juntos em uma mesma transação.	Determinar que produtos costumam ser colocados juntos em um carrinho de supermercado.
SEGMENTAÇÃO	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos.	Agrupar clientes por região do país Agrupar clientes com comportamento de compra similar Agrupar seções de usuários <i>web</i> para prever comportamento futuro de usuário.
SUMARIZAÇÃO	Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados.	Tabular o significado e desvios padrão para todos os itens de dados Derivar regras de síntese.

Tabela 1 – Tarefas realizadas por técnicas de mineração de dados

Fonte: Adaptado de (DIAS, 2002).

Na Tabela 1, encontra-se as cinco principais tarefas realizadas por técnicas de mineração com suas descrições e exemplos. Assim podendo ajudar a saber quais as tarefas que podem ser utilizadas conforme cada tipo de base de dados.

2.4 Bibliotecas para Mineração de Dados em Python

2.4.1 Pandas

Pandas é uma biblioteca de *software* escrita para a linguagem de programação *Python* para manipulação e análise de dados. Em particular, oferece estruturas de dados e operações para manipular tabelas numéricas e séries temporais².

2.4.2 Numpy

Numpy é uma biblioteca *Python* que é usada principalmente para realizar cálculos em *Arrays* Multidimensionais. O *NumPy* fornece um grande conjunto de funções e operações de biblioteca que ajudam os programadores a executar facilmente 38 cálculos numéricos. Esta biblioteca já vem importada dentro do *scikit-learn*, onde é utilizada³.

2.4.3 SciPy

Scipy é uma biblioteca baseado no *NumPy* e, portanto, estende seus recursos. A principal estrutura de dados *SciPy* é novamente uma matriz multidimensional, implementada pelo *Numpy*. O pacote contém ferramentas que ajudam a resolver álgebra linear, teoria da probabilidade, cálculo integral e muitas outras tarefas⁴.

2.5 Bibliotecas para Visualizações Gráficas em Python

2.5.1 Matplotlib

Matplotlib é uma biblioteca para *Python* para plotagem 2D, podendo gerar histogramas, gráficos, gráficos de barras, gráficos de erros, gráficos de dispersão de uma forma muito simples⁵.

2.5.2 Plotly

Plotly é uma biblioteca popular que permite construir facilmente gráficos sofisticados. O pacote é adaptado para trabalhar em aplicativos da *web* interativos. Entre suas visualizações notáveis estão gráficos de contorno, gráficos ternários e gráficos 3D⁶.

² Disponível em: <<https://pandas.pydata.org/>>

³ Disponível em: <<https://numpy.org/>>

⁴ Disponível em: <<https://www.scipy.org/>>

⁵ Disponível em: <<https://matplotlib.org/>>

⁶ Disponível em: <<https://plotly.com/>>

2.5.3 Seaborn

Seaborn é uma biblioteca do *Python* para visualização de dados. Provê uma interface em alto nível, apresentando gráficos estatísticos atraentes e informativos em diferentes formas e configurações⁷.

2.6 Bibliotecas para Aprendizado de Máquina em Python

2.6.1 Scikit-learn

Scikit-learn, é uma biblioteca de aprendizado de máquina de código aberto para a linguagem de programação *Python*, também conhecida como *sklearn*, sua abreviação e nome do pacote no *Python*. Ela inclui vários algoritmos de classificação, regressão e agrupamento incluindo máquinas de vetores de suporte, *K-means* e algoritmos de *clustering*, entre outros. E é projetada para interagir com as bibliotecas *Python* numéricas e científicas, *NumPy* e *SciPy*⁸.

2.6.2 K-Mens

É um algoritmo de agrupamento (*clustering*), com a abordagem de particionamento, que divide os dados em grupos e onde cada dado vai participar de apenas um grupo. É um algoritmo simples e muito utilizado nesta técnica. Primeiro é definido o K, algo informado pelo usuário, que é o número de *cluster* que o algoritmo irá agrupar. Cada *cluster* estará associado a um ponto central, o centroide. Então cada ponto do conjunto de dados de treinamento, será associado a um centroide, em qual está mais próximo. O centroide de cada *cluster* é atualizado com base nos pontos atribuídos ao *cluster*. Repete-se o processo de atribuição e atualização das etapas até que nenhum ponto altere os *clusters*, ou até que os centroides não se alterem, assim definindo os *clusters* conforme o K passado no início⁹.

2.7 Tribunal de Contas - TCE

O art. 70 da Constituição Federal de 1988 (BRASIL, 2009), determina que o exame das contas dos entes públicos brasileiros deve ser realizado pelo controle interno e externo. O controle externo fica sob responsabilidade do Poder Legislativo, com auxílio dos Tribunais de Contas, ao passo que o controle interno será levado a efeito por sistema do próprio ente público (CARVALHO, 2002).

O controle externo é exercido por um órgão que não faça parte da estrutura a qual pertencem os atos que devem ser controladas, como o controle que o Poder Judiciário

⁷ Disponível em: <<https://seaborn.pydata.org/>>

⁸ Disponível em: <<https://scikit-learn.org/stable/>>

⁹ Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>>

pode vir a realizar sobre atos do Poder público, o controle direto ou indireto a cargo do Congresso Nacional e o realizado pelo Tribunal de Contas (NETO, 2004).

De acordo com (TCU, 2020), os Tribunais de Contas dos estados, tem o dever de fiscalizar a parte contábil, financeira, operacional e patrimonial do uso dos recursos públicos. Portanto é possível perceber a necessidade de um mecanismo eficaz para detecção e correção da utilização dos recursos públicos de forma incorreta, tendo em vista a dimensão dos Tribunais de Contas.

De acordo com (TCE-TO, 2009), os Tribunais de Contas tem como função, acompanhar a gestão, fiscalizar e analisar os processos de prestações de contas, atos de admissão de pessoal, aposentadorias e pensões, licitações, contratos e convênios. No Tribunal de Contas, os dados técnicos apresentados, ou seja, os demonstrativos contábeis deverão necessariamente ser tabulados, analisados e confrontados com a legislação regente e emitido parecer sobre sua aprovação ou desaprovação (MELO, 2005).

Segundo (SANTANA, 2008), os Tribunais de Contas, atualmente, possuem as seguintes atribuições que estão estabelecidas na nossa Constituição Federal nos artigos 70 a 75:

- Apreciar as contas do chefe do Executivo;
- Julgar as contas dos Administradores Públicos e demais responsáveis por dinheiro, bens e valores públicos da administração direta e indireta;
- Apreciar, para fins de registro, a legalidade da contratação de pessoal, concessões de aposentadoria, reforma e revisão;
- Realizar auditoria e inspeções de natureza contábil, financeira, orçamentária, operacional e patrimonial nas entidades da administração direta e indireta, quanto à legalidade, legitimidade, economicidade, aplicação das subvenções e renúncia de receitas;
- Fiscalizar a aplicação de recursos públicos repassados;
- Aplicar sanções em caso de irregularidades;
- Determinar a adoção de providências para o exato cumprimento da Lei;
- Sustar ato impugnado se não atendido;
- Representar ao Poder competente sobre irregularidades ou abusos apurados.

(FIGUEIREDO, 2002) complementa que, o cumprimento da Lei de Responsabilidade Fiscal (LRF) deve ser verificado pelos Tribunais de Contas. A LRF está erguida sobre alguns pilares, entre eles está o da transparência. O objetivo e princípio da transparência

e permitir e incluir a participação social da população, assim podendo ser considerado uma forma mais eficaz na vigilância dos gestores.

3 Metodologia

A metodologia utilizada nesse trabalho foi uma pesquisa descritiva, com um estudo quantitativo. Foi realizada uma consulta em um banco de dados (*postgre*) através da aplicação web *jupyter* que dá suporte a diversas bibliotecas e linguagens de programação, incluindo *python*.

3.1 Materiais Utilizados

Para provimento do cenário foram utilizados os seguintes materiais descritos na Tabela 2.

DESCRIÇÃO	VERSÃO/MODELO
Sistema Operacional Linux	Ubuntu 18.04.5 LTS
Docker	19.03.6
PostgreSQL	13.1
Python	3.9
Jupyter	6.1.5
Matplotlib	3.3.3
Sklearn	0.23.2
Pandas	1.1.4
Numpy	1.19.4

Tabela 2 – Recursos/ferramentas utilizados
Fonte: Autoria própria.

Para a preparação do cenário, foi utilizado o sistema operacional *Linux Ubuntu* como terminal para acesso às informações armazenadas no banco *postgre* do Projeto TCE/UNITINS através da plataforma *jupyter* hospedado em um *docker*. Para a manipulação, análise e plotagem dos resultados foi necessário a instalação das seguintes bibliotecas mencionadas acima na Tabela 2.

3.2 Coleta dos Dados

Para obter os dados sobre despesas nos municípios do estado do Tocantins, foi utilizado uma base de dados gerada através de *scripts* de raspagens desenvolvido pelo projeto TCE/UNITINS, que coletam informações dos portais de transparência dos mesmos e armazenados em um banco de dados *postgre*. Esses dados de despesas são disponibilizados em formato de tabelas como apresentado na Tabela 3, e podem ser acessados através do *jupyter*.

dt_emissao	numero	valor	servico_produto	organizacao_id
09/11/2020	53932	2341.25	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53945	2200	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53939	1800.8	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53941	2250	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53942	1045	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53931	2338.72	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53934	2517.74	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53933	2341.25	LIQUIDACAO DE EMPENHO...	1
09/11/2020	53940	2091.51	LIQUIDACAO DE EMPENHO...	1
06/11/2020	53870	994.19	LIQUIDACAO DE EMPENHO...	1
05/11/2020	53936	1045	LIQUIDACAO DE EMPENHO...	1
04/11/2020	53854	375.05	LIQUIDACAO DE EMPENHO...	1
04/11/2020	53871	500	LIQUIDACAO DE EMPENHO...	1
03/11/2020	53822	7000	LIQUIDACAO DE EMPENHO...	1
03/11/2020	53823	2200	LIQUIDACAO DE EMPENHO...	1

Tabela 3 – Exemplo de campos utilizados na análise de dados
Fonte: Base de Dados TCE/UNITINS.

Na tabela possui vários campos que a compõe como: "id", "empenho", "dt_emissao", "dt_create", "numero", "valor", "serviço_produto", "nota", e "organização_id". Como o objetivo desse estudo está relacionado com os processo de despesas, não poderiam deixar de fazer parte da seleção dos dados as colunas, "dt_emissao", "numero", "valor", "serviço_produto", e "organização_id" que pode ser representado por "1" para Prefeitura ou "2" para Câmara Municipal.

3.3 Análise dos Dados

A ferramenta utilizada para a execução do processo foi o *jupyter*¹⁰. É um *software open-source* com interface gráfica *web*, iterativo e fácil manuseio, permite varias linguagens de programação, e além de permitir plotagem de gráficos e *script*.

Como descrito anteriormente, existem vários algoritmos e bibliotecas que podem ser utilizadas para encontrar padrão em uma base de dados, mais especialmente nesse trabalho na base de dados do Projeto TCE/UNITINS do Tribunal de Contas do Estado do Tocantins. Tendo como objetivo encontrar padrões relevantes com indícios de gastos municipais com valores elevados em um determinado período para municípios do Tocantins. Com intuito de encontrar características semelhantes entre os municípios, foi escolhido então o algoritmo de agrupamento de dados *k-means*. Para agrupar os municípios semelhantes, foram utilizados alguns critérios mencionados abaixo:

¹⁰ Disponível em: <<https://jupyter.org/>>

- PIB;
- População.

Todos os dados se encontram disponíveis no IBGE. Após definir os critérios para agrupamento, como resultado foram gerados 5 grupos entre os 139 municípios que apresentam características semelhantes entre os que serão utilizados para a análise de dados.

Nesse contexto, todos os dados de despesas municipais são enviados para o portal transparência de cada município. Dessa forma o projeto TCE/UNITINS fez a coleta dessas informações dos portais transparência dos Municípios do Tocantins através de *scripts* de raspagem como mencionado anteriormente, essas informações são armazenados em um banco de dados em forma de dados brutos, sendo necessário fazer uma transformação desses dados para que seja utilizado somente informações que sejam importante para o desenvolvimento do projeto. Dessa forma, a plataforma *jupyter* possibilitou fazer a manipulação desses dados através do console *Django Shell* disponível na plataforma. A informações do banco foi acessada através de *models* e armazenadas em *dataframe*. Dando uma origem a um novo grupo de informações já filtradas.

3.4 Procedimento Realizado

Após a coleta dos dados brutos, é preciso fazer uma seleção mais criteriosa antes que seja feita a análise e interpretação dos mesmos, para que os resultados tenham um nível de confiança maior. Portanto, foram selecionados os dados de despesas disponíveis na base TCE/UNITINS entre o período de setembro de 2019 a setembro de 2020, uma amplitude temporal de 13 meses, pelo fato de ser uma base de dados com mais de 3GB de informações armazenadas e em crescente andamento, é necessário um poder computacional maior para que seja utilizada todas as informações armazenadas, desde modo é justificado a escolha de somente 13 meses para análise. Também foi removido alguns colunas desnecessárias para a descoberta de conhecimento como, "id", "empenho", "dt_create", e "nota".

Foi também agregado as despesas mensais de municípios pertencente ao grupo, encontrando assim valores esperados dos grupos, utilizando algoritmo de regressão linear, através do processo de equação para se estimar o valor esperado, dados os valores de algumas outras variáveis x (despesas por município). Dessa forma, foi possível comparar as despesas efetivadas pelo município, com o valor esperado do grupo pertencente a esse município analisado, assim identificando anomalias quando o valor da liquidação ultrapassar o valor esperado no gráfico.

Por se tratar de grandezas diferentes de valores (despesa municipal e valor esperado do grupo) foi feito ao processo de padronização para que os dados tenha uma faixa de mesma escala, assim dando a possibilidade de ser comparado. Essa técnica na prática

ignora a forma da distribuição e transforma os dados para forma com média próxima de zero e um desvio padrão próximo a um, ou seja, assume que não temos valores discrepantes nos dados. A Figura 2 ilustra todo o processo de coleta, preparação e análise dos dados.

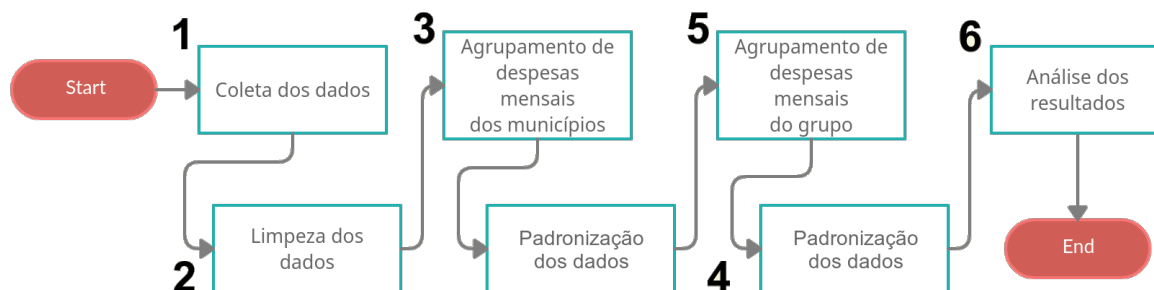


Figura 2 – Pipeline Etapas de Preparação dos Dados

Fonte: Autoria própria.

As etapas abaixo especifica-se a seguir o que cada processo significa:

- Etapa 1: Coletar todas as informações do banco de dados *postgre* através de *models* e armazenar em um *dataframe* com amplitude temporal de 13 meses;
- Etapa 2: Remoção de alguns colunas desnecessárias como, "id", "empenho", "dt_create", e "nota";
- Etapa 3: Agregação de todas as despesas do município por meses;
- Etapa 4: Padronização dos dados para a mesma escala de grandeza;
- Etapa 5: Agregação das despesas mensais dos municípios pertencente ao grupo;
- Etapa 6: Plotagem dos resultados.

Na Tabela 4 é possível observar os valores da coluna antes de passar pelo processo de padronização, e já na Tabela 5 é possível observar a nova coluna após o processo de padronização.

09-19	813351.98
10-19	915142.98
11-19	847181.34
12-19	1523810.00
01-20	1221915.00
02-20	960733.06
03-20	778211.15
04-20	866449.20
05-20	992766.31
06-20	978849.11
07-20	1541222.76
08-20	866240.82
09-20	737994.14

Tabela 4 – Coluna antes do processo de padronização
Fonte: Autoria própria.

09-19	-0.748706
10-19	-0.347640
11-19	-0.615415
12-19	2.050564
01-20	0.861070
02-20	-0.168011
03-20	-0.887164
04-20	-0.539498
05-20	-0.041797
06-20	-0.096632
07-20	2.119172
08-20	-0.540319
09-20	-1.045623

Tabela 5 – Coluna após o processo de padronização
Fonte: Autoria própria.

Como pode ser observado os dados ficaram todos padronizados e o desvio padrão está em torno de 1. Existem outros tipos de métodos de padronização e usar cada um é uma escolha que deve tomar baseado nos dados que serão utilizados.

4 Resultados

Nesse tópico, serão apresentados os resultados obtidos a partir do conjunto de etapas definidos no capítulo 3. Esta é a fase de interpretação e validação dos padrões que foram extraídos do processo de Descoberta de Conhecimento em Base de Dados.

4.1 Clusterização dos Municípios

Foram utilizados alguns filtros para determinar o escopo para geração dos *clusters* dos municípios como mencionado no capítulo 3, foi escolhido o PIB e população para que seja o filtro de agregação dos municípios semelhantes. Dessa forma os dados serão mais aproveitados e os resultados mais satisfatórios. Para atingir esse objetivo, foi aplicado na base extraída do IBGE o algoritmo de *k-means* tendo como parâmetro a criação de 5 grupos para os municípios semelhantes, resultando então os *clusters* como apresentado na Figura 3 ou os municípios detalhados na Tabela 6.

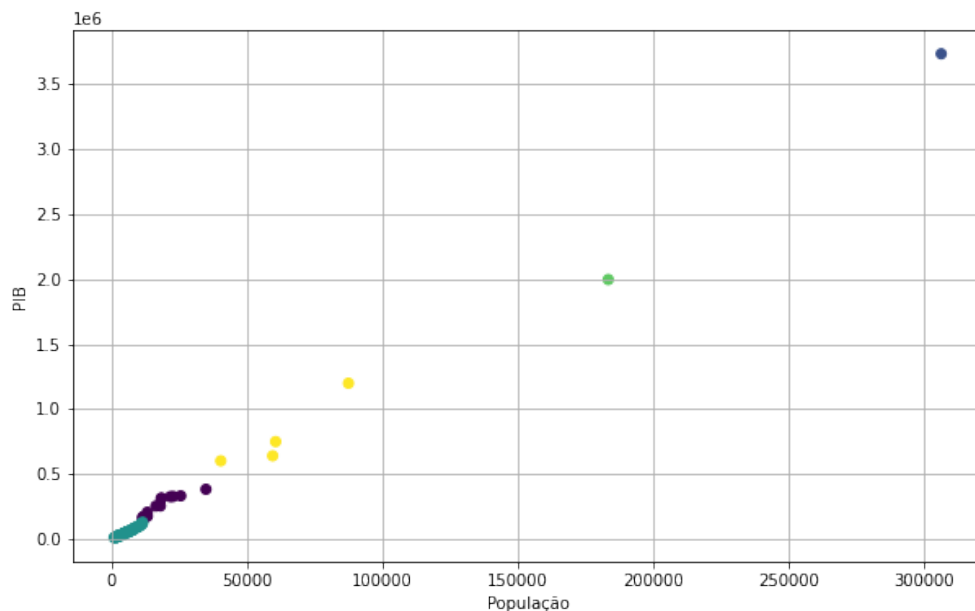


Figura 3 – Divisão dos Municípios do Tocantins em 5 Clusters

Fonte: Autoria própria.

GRUPOS	MUNICÍPIOS
GRUPO 0	Abreulândia, Aguiarnópolis, Aliança do Tocantins, Almas, Ananás, Angico, Aparecida do Rio Negro, Aragominas, Araguacema, Araguanã, Araguaçu, Arapoema, Arraias, Augustinópolis, Aurora do Tocantins, Axixá do Tocantins, Babaçulândia, Bandeirantes do Tocantins, Barra do Ouro, Barrolândia, Bernardo Sayão, Bom Jesus do Tocantins, Brasilândia do Tocantins, Brejinho de Nazaré, Buriti do Tocantins, Cachoeirinha, Cariri do Tocantins, Carmolândia, Carrasco Bonito, Caseara, Centenário, Chapada da Natividade, Chapada de Areia, Colméia, Combinado, Conceição do Tocantins, Couto Magalhães, Cristalândia, Crixás do Tocantins, Darcinópolis, Divinópolis do Tocantins, Dois Irmãos do Tocantins, Dueré, Esperantina, Fátima, Figueirópolis, Filadélfia, Fortaleza do Tabocão, Goianorte, Ipueiras, Itacajá, Itaguatins, Itapiratins, Itaporã do Tocantins, Jaú do Tocantins, Juarina, Lagoa do Tocantins, Lajeado, Lavandeira, Lizarda, Luzinópolis, Marianópolis do Tocantins, Mateiros Goiatins, Maurilândia do Tocantins, Miranorte, Monte do Carmo, Monte Santo do Tocantins, Muricilândia, Natividade, Nazaré, Nova Rosalândia, Novo Acordo, Novo Alegre, Novo Jardim, Oliveira de Fátima, Palmeirante, Palmeiras do Tocantins, Palmeirópolis, Pau D'Arco, Pequizeiro, Pindorama do Tocantins, Piraquê, Pium, Ponte Alta do Bom Jesus, Ponte Alta do Tocantins, Porto Alegre do Tocantins, Praia Norte, Presidente Kennedy, Pugmil, Recursolândia, Riachinho, Rio da Conceição, Rio dos Bois, Rio Sono, Sampaio, Sandolândia, Santa Fé do Araguaia, Santa Maria do Tocantins, Santa Rita do Tocantins, Santa Rosa do Tocantins, Santa Tereza do Tocantins, Santa Terezinha do Tocantins, São Bento do Tocantins, São Félix do Tocantins, São Miguel do Tocantins, São Salvador do Tocantins, São Sebastião do Tocantins, São Valério, Silvanópolis, Sítio Novo do Tocantins, Sucupira, Taipas do Tocantins, Talismã, Tocantínia, Tupirama, Tupiratins, Wanderlândia
GRUPO 1	Araguaína.
GRUPO 2	Palmas.
GRUPO 3	Gurupi, Miracema do Tocantins, Paraíso do Tocantins, Porto Nacional
GRUPO 4	Alvorada, Araguatins, Campos Lindos, Colinas do Tocantins, Dianópolis, Formoso do Araguaia, Guaraí, Lagoa da Confusão, Nova Olinda, Paranã, Pedro Afonso, Peixe, Taguatinga, Tocantinópolis, Xambioá.

Tabela 6 – Separação dos municípios por grupos
Fonte: Autoria própria.

Com os municípios agrupados por *clusters*, garante a possibilidade de trabalhar a comparação entre eles já que possuem características semelhantes, e calcular o valor esperado para o grupo que o município analisado pertence. Assim verificando as anomalias detectadas.

4.2 Comparação dos Municípios

Para que seja feita a comparação do município com o valor esperado do grupo que o mesmo pertence, foram utilizados os valores mensais das despesas já padronizados do município através do algoritmo de padronização e comparado com o valores mensais esperado para o grupo pertencente ao município com os dados também padronizados.

4.2.1 Análise Município de Rio da Conceição

Para melhor visualização e entendimento da análise do Município Rio da Conceição foi utilizado a biblioteca *matplotlib* para geração de gráficos como apresentados na imagem a seguir.

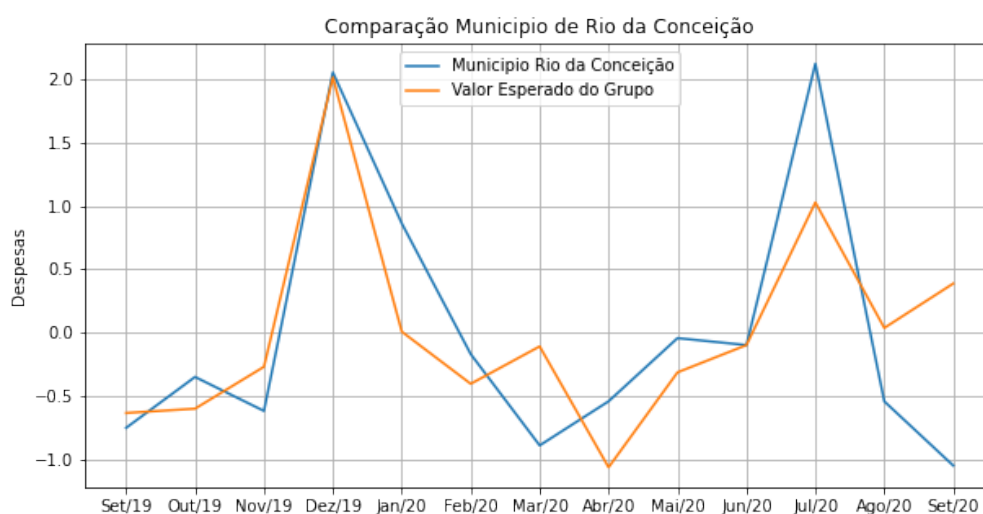


Figura 4 – Comparação município de Rio da Conceição

Fonte: Autoria própria.

Na Figura 4 é possível observar que após feita a comparação do Município de Rio da Conceição com o valor esperado do seu grupo, os meses de out/2019, dez/2019, jan/2020, fev/2020, abr/2020, mai/2020 e jul/2020 as despesas municipais foram mais altas que o valor esperado, onde a linha azul representa os gastos municipais e a linha amarela o valor esperado do grupo. Na Tabela 7 retirada dos dados originais, é possível verificar os possíveis lançamentos que podem ter contribuído para essa anomalia no mês de julho de 2020, onde as despesas extrapolam o valor esperado.

empenho	liquidacao	numero	valor	historico
19032	20413	24217	85000	PAGAMENTO REF. A AQUISIÇÃO DE 1 VEÍCULO AMBULÂNCIA...
19032	20412	24216	85000	PAGAMENTO REF. AQUISIÇÃO DE 1 VEÍCULO AMBULÂNCIA...
19482	19626	23378	70046	VALOR REF: VENCIMENTOS DOS SERVIDORES DO FUNDO...
24331	20514	24331	46000	AO RESTANTE PAGAMENTO DA PRESTAÇÃO DE SERVIÇOS...
19682	19829	23589	30000	VALOR REFERENTE: DESPESA COM FORNECIMENTO DE MATERIAIS DE ILUMINAÇÃO...

Tabela 7 – Maiores lançamentos para mês de Jul/2020 - Rio da Conceição
Fonte: Autoria própria.

4.2.2 Análise Município de Centenário

Na figura a seguir também foi feita a projeção de análise do município de Centenário com objetivo de identificar anomalias pelo periodo coletado.

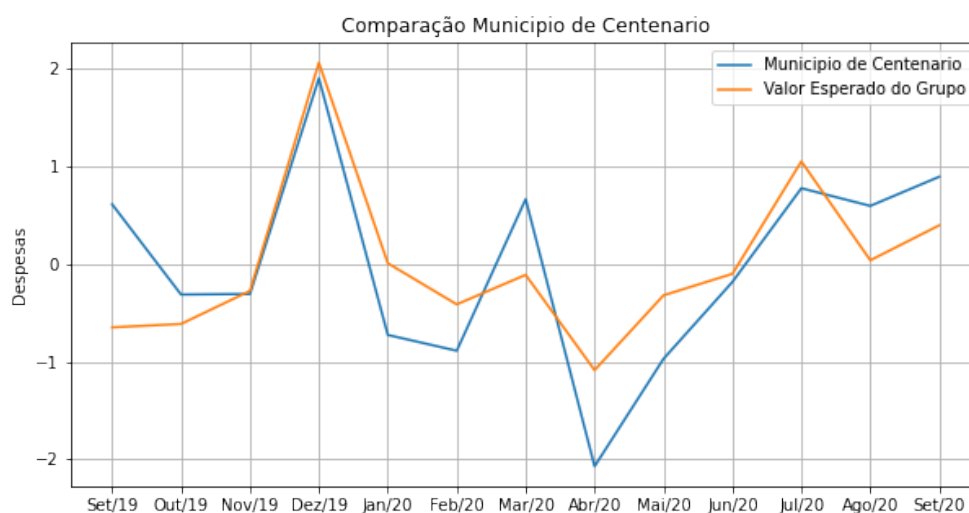


Figura 5 – Comparação município de Centenário
Fonte: Autoria própria.

Na Figura 5 também é possível observar que entre os meses de set/2019, mar/2020, ago/2020 e set/2020 ultrapassaram o valor esperado do grupo. Dessa forma na Tabela 8 é possível encontrar os possíveis lançamentos que podem ter contribuído para essa anomalia para o município de Centenário no mês de setembro de 2019.

empenho	liquidacao	numero	valor	historico
20739	25916	29685	115276.62	PAGAMENTO DE DESPESA PRESTAÇÃO DE SERVIÇOS...
20799	26027	29907	71371.31	Valor que se empenha para atender à despesa com VENCIMENTOS...
20797	26025	29905	37273.35	Valor que se empenha para atender à despesa com VENCIMENTOS...
18894	26002	29823	37203.7	PAGAMENTO DE DESPESA COM PRESTAÇÃO DE SERVIÇOS...
20793	26021	29968	26941.29	PAGAMENTO DE DESPESA COM CONTRIBUIÇÃO PATRONAL-INSS...

Tabela 8 – Maiores lançamentos para mês de Set/2019 - Centenário
Fonte: Autoria própria.

4.2.3 Análise Município de Cachoeirinha

Da mesma forma foi analisado o município de Cachoeirinha com intuito de identificar anomalias como apresentado na figura abaixo.

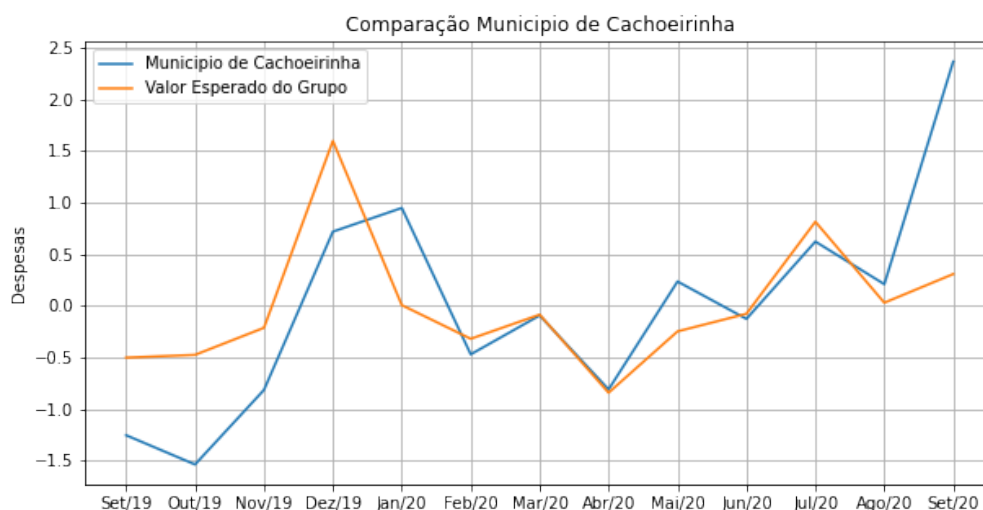


Figura 6 – Comparação município de Cachoeirinha
Fonte: Autoria própria.

Na Figura 6 observamos que os meses jan/2020, abr/2020, mai/2020, ago/2020 e set/2020 as despesas municipais ultrapassaram o valor esperado do grupo. Na Tabela 9 verifica-se os lançamentos mais altos com possibilidade de ter contribuído com essa anomalia para o mês analisado.

empenho	liquidacao	numero	valor	historico
12964	8727	15622	192690.16	RECOLHIMENTO DO INSS PARTE PATRONAL.
12963	8725	15618	106488.84	RECOLHIMENTO DE PREVIDÊNCIA PARTE PATRONAL...
12964	8727	15623	99668.35	RECOLHIMENTO DO INSS PARTE PATRONAL.
12964	8727	15624	90229.59	RECOLHIMENTO DO INSS PARTE PATRONAL.
12028	8645	15527	82979.4	VENCIMENTOS DEVIDOS AOS PROFESSORES LOTADOS...

Tabela 9 – Maiores lançamentos para mês de Set/2020 - Cachoeirinha
Fonte: Autoria própria.

4.2.4 Análise Município de Santa Terezinha

Na figura a seguir podemos observar a comparação do município de Santa Terezinha com o valor esperado do grupo.

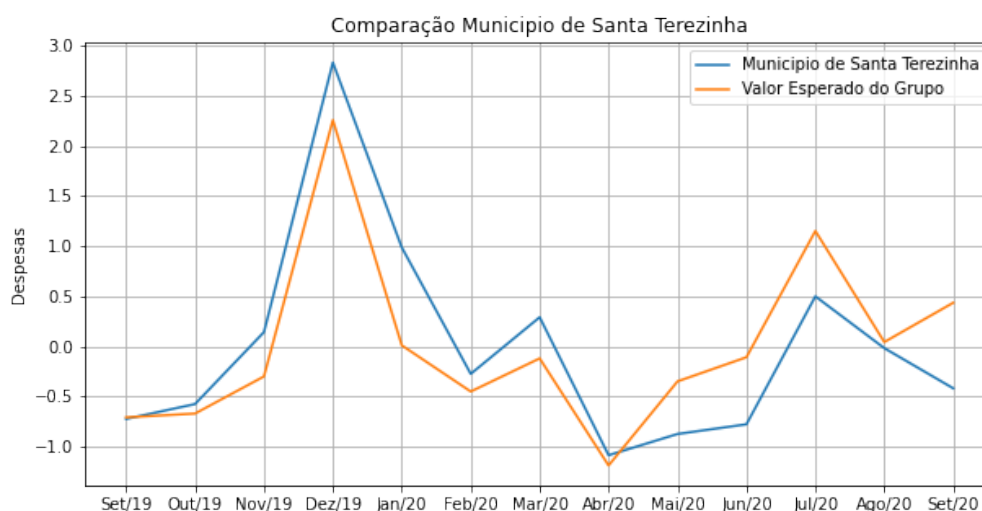


Figura 7 – Comparação município de Santa Terezinha
Fonte: Autoria própria.

Na Figura 7 observa-se que os meses de out/2019, nov/2019, dez/2019, jan/2020, fev/2020, mar/2020 e abr/2020 as despesas municipais foram superior ao valor esperado. Dessa forma verifica-se na Tabela 10 os maiores lançamentos para o mês de dezembro de 2019 do município de Santa Terezinha.

empenho	liquidacao	numero	valor	historico
23870	32377	35566	269610	AJUSTE DA CONSOLIDAÇÃO CÂMARA MUNICIPAL DE SANTA...
23390	31315	34429	79523.18	PAGAMENTO DE DESPESA REFERENTE AOS VENCIMENTOS...
23902	32409	35598	70000	AJUSTE DA CONSOLIDAÇÃO CÂMARA MUNICIPAL DE SANTA...
23871	32378	35567	64285.31	AJUSTE DA CONSOLIDAÇÃO CÂMARA MUNICIPAL DE SANTA...
22421	31312	34425	48819.66	PAGAMENTO DE DESPESA REFERENTE AOS VENCIMENTOS...

Tabela 10 – Maiores lançamentos para mês de Dez/2019 - Santa Terezinha

Fonte: Autoria própria.

Desse modo, foi possível realizar a comparação de alguns municípios do grupo 0 e apresentar alguns dos maiores lançamentos para o determinado mês analisado. A escolha dos municípios apresentados acima foram devido à apresentarem valores de despesas muito acima do esperado, quando comparado ao valor esperado do grupo pertencente. Mas também à possibilidade de analisar outros municípios de outros grupo, respeitando todo o processo do *pipeline* disponível na Figura 2.

5 Conclusão

A finalidade desse trabalho consistiu em aplicar técnicas de mineração de dados para o processo de descoberta de conhecimento em uma base do Projeto TCE/UNITINS, efetuando uma série de etapas, que vão desde o processo de coleta, preparação, limpeza dos dados e extração dos padrões, até chegar a avaliação. Onde é analisado o quanto os padrões identificados agregam valor em relação ao problema em questão.

Em relação a esse trabalho, é possível compreender como os recursos de *data analytics* podem sim ser essencial no apoio de fiscalização. Principalmente se a empresa ou organização pertence a um ambiente dinâmico, onde se encontra em constantes mudanças. Ainda é importante ressaltar que a utilização desses recursos auxiliam também no momento de cumprimento da lei em relação a transparência do sistema de fiscalização de gestão pública.

Para chegar ao final desse trabalho, foi necessário alcançar alguns objetivos específicos como: analisar os conceitos e técnicas de *data mining*, verificar a aplicabilidade das técnicas de *data mining* para o processo de análise na base TCE/UNITINS e gerar bases com municípios que contém características semelhantes para comparação.

Desse forma o projeto cumpre a função de identificar os municípios que apresentam despesas municipais com valores acima do valor esperado. Entretanto de forma alguma significa afirmar que é uma irregularidade e sim uma ferramenta de auxílio e apoio à auditoria. Portanto é possível a utilização da *data mining* para auxiliar auditores na fiscalização trazendo mais transparência, eficiência e eficácia aos processos analisados neste âmbito.

Na seção seguinte são apresentadas os possíveis trabalhos futuros.

5.1 Trabalhos Futuros

Pretende-se como trabalhos futuros:

- Geração de alertas à auditoria de forma automática mensal por município;
- Criação de painel para acesso interativo dos auditores;
- Expansão para novas áreas de auditoria.

Referências

- ARONSON, E. T. R. S. J. *Business Intelligence: um enfoque gerencial para a inteligência do negócio*. [S.l.]: Porto Alegre, 2009.
- BATISTA, J. *Estruturando um Data Lake para Transformação Digital*. [S.l.]: Disponível em: <<https://visagio.com/pt/insights/estruturando-um-data-lake-para-transformacao-digital>> . Acesso em: 21 dezembro. 2020, 2017.
- BITENCOURT, R. *Controle da transparência na contratação pública no Brasil – o acesso à informação como forma de viabilizar o controle social da Administração Pública*. [S.l.]: Disponível em: <<http://dx.doi.org/10.17058/rdunisc.v2i49.7892>> . Acesso em: 21 dezembro. 2020, 2016.
- CARVALHO, D. *Data Mining: Definição, importância e aplicação na gestão organizacional*. [S.l.: s.n.], 2002.
- CARVALHO, L. A. V. de. *Data Mining – A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. [S.l.]: Revista Eletrônica Gestão e Serviços, 2015.
- COSTA, E. *Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações*. [S.l.]: Jornada de Atualização em Informática na Educação, 2012.
- COSTA, R. *Data mining: um guia prático*. [S.l.]: Rio de Janeiro: Elsevier, 2005.
- DIAS, M. M. *Parâmetros na escolha de técnicas e ferramentas de mineração de dados*. [S.l.]: Acta Scientiarum. Technology, 2002.
- FAYYAD, U. *From Data Mining to Knowledge Discovery in Databases*. [S.l.]: American Association for Artificial Intelligence, 1996.
- FIGUEIREDO, C. M. *Ética na gestão pública e exercício da cidadania: o papel dos tribunais de contas brasileiros como agências de accountability - O caso do Tribunal de Contas de Pernambuco*. [S.l.]: VII Congresso Internacional del CLAD, Lisboa - Portugal, 2002.
- FIGUEIREDO, R. M. da C. *Governança em Tecnologia de Informação e Comunicação para o Setor Público*. [S.l.]: Universidade de Brasília – UnB, 2018.
- FIGUEIREDO, V. da S. *Transparência e controle social na administração pública*. [S.l.]: Disponível em: <<https://www.fclar.unesp.br/Home/Departamentos/AdministracaoPublica/RevistaTemasdeAdministracaoPublica/vanuza-da-silva-figueiredo.pdf/>> . Acesso em: 24 agosto. 2020, 2017.
- FREITAS, O. G. J. *Um modelo de Sistema de Gestão do Conhecimento para Grupos de Pesquisa e Desenvolvimento*. [S.l.]: San Francisco: Morgan Kaufmann Publishers, 2003.
- GOEBEL, D. *Data mining no contexto de customer relationship management*. [S.l.]: Disponível em: <<http://www.ead.fea.usp.br/cad-pesq/arquivos/v12n2art6.pdf/>> . Acesso em: 14 agosto. 2020, 2018.

- HEEKS, L. G. R. *Measuring the Barriers to Big data for Development*. [S.l.]: Disponível em: <http://hummedia.manchester.ac.uk/institutes/gdi/publications/workingpapers/di/di_wp62.pdf>. Acesso em: 21 dezembro. 2020, 2016.
- HEUDECKER, N. *Big data. Gartner Insight, Gartner Research*. [S.l.]: Disponível em: <<https://www.gartner.com/doc/2626815/predicts-big-data/>> . Acesso em: 21 dezembro. 2020, 2014.
- KAMBER, J. H. M. *Data mining: concepts and techniques*. [S.l.]: Universidade Federal de Santa Catarina, 2006.
- KUMAR, V. *Introdução ao Data Mining Mineração de Dados*. [S.l.]: Ciencia Moderna - Rio de Janeiro, 2009.
- MEDEIROS, B. C. *Computação em Nuvem: uma análise bibliométrica dos estudos publicados em eventos e periódicos no Brasil*. [S.l.]: Botucatu, SP, Tekne e Logos, 2015.
- MELO, P. *TRIBUNAIS DE CONTAS NO BRASIL: natureza jurídica e alcance das suas funções, à luz dos princípios constitucionais da Administração Pública*. [S.l.]: Universidade do Vale do Itajaí - UNIVALI, 2005.
- NETO, J. *Sistemas de informação e as decisões gerenciais na era da internet*. [S.l.: s.n.], 2004.
- PASTA, A. *Aplicação da técnica de Data Mining na base de dados do ambiente de gestão educacional*. [S.l.]: Universidade do Vale do Itajaí, 2011.
- SANTANA, E. W. *Uma análise sobre a utilização da informação contábil como um instrumento de controle social dos municípios paraibanos com mais de 50.000 habitantes*. [S.l.]: Universidade Federal da Paraíba e da Universidade Federal do Rio Grande do Norte, 2008.
- SILVA, P. G. da. *Dificuldades de usuários para acessar informações de portais de transparência de estados brasileiros*. [S.l.]: Revista Gestão e Organizações ISSN 2526-2289 v. 03, n. 02, 2018.
- SOUZA, J. T. *Métodos de Seleção de Atributos e Análise de Componentes Principais*. [S.l.]: Universidade Tecnológica Federal do Paraná, 2017.
- TCE-TO. *Tribunal de Contas do Estado do Tocantins*. [S.l.]: Disponível em: <<https://www.tceto.tc.br/>> . Acesso em: 12 novembro. 2020, 2009.
- TCU. *Competências do TCU*. [S.l.]: Disponível em: <<https://portal.tcu.gov.br/institucional/conheca-o-tcu/competencias/>> . Acesso em: 23 dezembro. 2020, 2020.
- TIME, M. *O que é Data Science?* [S.l.]: Disponível em: <<http://blog.mjv.com.br/ideias/o-que-e-data-science>> . Acesso em: 21 dezembro. 2020, 2015.
- TREVISAN, M. G. *O uso da Mineração de Dados na descoberta de conhecimento em empresa do setor agrícola*. [S.l.]: Universidade de Araraquara. Araraquara-SP, 2017.
- VIEIRA, S. C. *Computação em Nuvem: Análise bibliométrica da produção científica sobre os fatores que influenciam as empresas no seu uso*. [S.l.]: Revista Eletrônica Gestão e Serviços, 2015.

VILARINHO, R. A. *Uso de Técnicas de Mineração de Dados para Classificação das Ocorrências de Casos de Dengue nos Municípios Brasileiros*. [S.l.]: Disponível em: <https://www.monografias.ufop.br/bitstream/35400000/326/1/MONOGRRAFIA_UsoTecnicaMineracao.pdf>. Acesso em: 24 agosto. 2020, 2018.

YAMAGUCHI, J. K. *Diretrizes para a escolha de técnicas de visualização aplicadas no processo de extração do conhecimento*. [S.l.]: Universidade Estadual de Maringá, Maringá, 2010.

ZANELLA, D. *Data Mining: Definição, importância e aplicação na gestão organizacional*. [S.l.]: Disponível em: <<https://www.yumpu.com/s/xDjzHnqVfrZnQOoF/>>. Acesso em: 26 agosto. 2020, 2017.