

**INSTITUTO FEDERAL GOIANO – CAMPUS CERES**  
**BACHARELADO EM SISTEMAS DE INFORMAÇÃO**  
**PAULO HENRIQUE RODRIGUES ARAUJO**

**USO DE MINERAÇÃO DE DADOS PARA AVALIAÇÃO DO DESEMPENHO**  
**DA JUSTIÇA ESTADUAL**

**CERES – GO**  
**2019**

**PAULO HENRIQUE RODRIGUES ARAUJO**

**USO DE MINERAÇÃO DE DADOS PARA AVALIAÇÃO DO DESEMPENHO  
DA JUSTIÇA ESTADUAL**

Trabalho de curso apresentado ao curso de Bacharelado em Sistemas de Informação do Instituto Federal Goiano – Campus Ceres, como requisito parcial para a obtenção do título de bacharel em Sistemas de Informação, sob orientação do Prof. Dr. Marcos de Moraes Sousa.

**CERES – GO**

**2019**



**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

**Identificação da Produção Técnico-Científica**

- |                                                                      |                                                         |
|----------------------------------------------------------------------|---------------------------------------------------------|
| <input type="checkbox"/> Tese                                        | <input checked="" type="checkbox"/> Artigo Científico   |
| <input type="checkbox"/> Dissertação                                 | <input type="checkbox"/> Capítulo de Livro              |
| <input type="checkbox"/> Monografia - Especialização                 | <input type="checkbox"/> Livro                          |
| <input type="checkbox"/> TCC - Graduação                             | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ |                                                         |

Nome Completo do Autor: Paulo Henrique Rodrigues Azeiteiro

Matricula: 2016103202020087

Título do Trabalho: Uso de mineração de dados para análise do desempenho de justiça estadual

**Restrições de Acesso ao Documento**

Documento confidencial:  Não  Sim, justifique: \_\_\_\_\_

Informe a data que poderá ser disponibilizado no RIIF Goiano: 29/10/13

O documento está sujeito a registro de patente?  Sim  Não

O documento pode vir a ser publicado como livro?  Sim  Não

**DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA**

O/A referido/a autor/a declara que:

- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Ceres \_\_\_\_\_ 29/10/13  
Local Data

Paulo Henrique Rodrigues Azeiteiro  
Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:

[Assinatura]  
Assinatura do(a) orientador(a)

Sistema desenvolvido pelo ICMC/USP  
Dados Internacionais de Catalogação na Publicação (CIP)  
**Sistema Integrado de Bibliotecas - Instituto Federal Goiano**

A663u Araujo, Paulo Henrique Rodrigues  
USO DE MINERAÇÃO DE DADOS PARA AVALIAÇÃO DO  
DESEMPENHO DA JUSTIÇA ESTADUAL / Paulo Henrique  
Rodrigues Araujo; orientador Marcos de Moraes Sousa. -  
- Ceres, 2019.  
28 p.

Monografia ( em Sistemas de Informação) --  
Instituto Federal Goiano, Campus Ceres, 2019.

1. Mineração de Dados. 2. Desempenho do  
Judiciário. 3. Justiça Estadual. I. Sousa, Marcos de  
Moraes, orient. II. Título.

#### ANEXO IV - ATA DE DEFESA DE TRABALHO DE CURSO

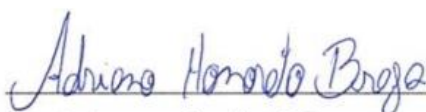
Ao(s) 22 dia(s) do mês de outubro do ano de dois mil e dezenove, realizou-se a defesa de Trabalho de Curso do(a) acadêmico(a) Paulo Henrique Rodrigues, do Curso de Bacharelado em Sistemas de Informação matrícula 2016103202030087, cujo título é "Uso de mineração de dados para avaliação do desempenho da justiça estadual". A defesa iniciou-se às 14 horas e 46 minutos, finalizando-se às 15 horas e 11 minutos. A banca examinadora considerou o trabalho aprovado com média 8,6 no trabalho escrito, média 8,6 no trabalho oral, apresentando assim média aritmética final 8,6 de pontos, estando o(a) estudante apto para fins de conclusão do Trabalho de Curso.

Após atender às considerações da banca e respeitando o prazo disposto em calendário acadêmico, o(a) estudante deverá fazer a submissão da versão corrigida em formato digital (.pdf) no Repositório Institucional do IF Goiano – RIIF, acompanhado do Termo Ciência e Autorização Eletrônico (TCAE), devidamente assinado pelo autor e orientador.

Os integrantes da banca examinadora assinam a presente.


  
Assinatura Presidente da Banca

MARCOS DE MORAES SOUZA



Assinatura Membro 1 Banca Examinadora

Adriano Honorato Braga

  
Assinatura Membro 2 Banca Examinadora

Leonardo Paulo Arantes

## **AGRADECIMENTOS**

A Deus, por ter me dado forças e sabedoria para superar toda e qualquer dificuldade encontrada.

Ao Instituto Federal Goiano – *Campus Ceres*, seu corpo docente, direção e administração pela oportunidade de concluir mais essa jornada em minha vida.

Ao meu orientador, professor Marcos de Moraes, pelo suporte no tempo que lhe coube, pelas suas correções, paciência e compreensão.

Agradeço a minha mãe, minha namorada e meus amigos, pelos incentivos e apoio.

Agradeço a todos que diretamente ou indiretamente me ajudaram para realização deste trabalho e elaboração deste documento, deixo aqui meu muito obrigado!

## RESUMO

A avaliação do desempenho do judiciário mensura o quão produtivo são os tribunais. As técnicas de mineração de dados permitem encontrar padrões e anomalias dado um conjunto de dados. Neste trabalho analisou-se a produtividade dos tribunais da justiça estadual brasileira por meio de técnicas de mineração de dados e avaliou-se o desempenho dos algoritmos de mineração de dados. O conselho nacional de justiça disponibiliza anualmente dados sobre a gestão judiciária. Estudos revisados apontam as principais variáveis utilizadas para avaliar o desempenho do judiciário, a quantidade de sentenças proferidas, dentre outras, é a principal variável apontada como critério para avaliar a produtividade dos tribunais. Também foi abordado o processo de extração de conhecimento em banco de dados, o *Knowledge Discovery in Databases*, ou KDD. Para avaliar a produtividade do judiciário tratou-se a quantidade de sentenças proferidas por magistrado, também se considerou a quantidade de casos novos dividido pela quantidade de sentenças proferidas por magistrado para refletir sobre o congestionamento dos tribunais. Para fins de comparações analisou-se os dados entre 2009 e 2017, agrupando-os por estado e região brasileira. O desempenho dos algoritmos foi mensurado através do teste t. Resultados obtidos na pesquisa apontaram os tribunais estaduais mais produtivos, como o estado do Rio de Janeiro que tem um grande número de sentenças e o estado do Pará que tem um dos menores números de congestionamento. A região nordeste apresenta o maior número de sentenças por magistrados e o centro-oeste tem o menor congestionamento. Apesar da dificuldade e falta de metodologia científica para mensurar o desempenho do judiciário as metodologias abordadas permitem boas comparações e uma discussão abrangente.

**Palavras-chave:** Mineração de dados, Desempenho do Judiciário, Justiça Estadual.

## ABSTRACT

Judicial performance evaluation measures how productive courts are. Data mining techniques allow finding patterns and anomalies given a data set. In this paper we analyzed the productivity of Brazilian state courts through data mining techniques and evaluated the performance of data mining algorithms. The national council of justice annually makes available data on judicial management. Revised studies indicate the main variables used to evaluate the performance of the judiciary, the number of sentences handed down, among others, is the main variable identified as a criterion to evaluate the productivity. Also discussed was the process of extracting knowledge from the database, the Knowledge Discovery in Databases, or KDD. To assess the productivity of the judiciary, the number of judgments handed down by the magistrate was treated, the number of new cases divided by the number of judgments handed down by the magistrate to reflect on the congestion of the courts. For comparison purposes, data were analyzed between 2009 and 2017, grouping them by state and Brazilian region. The performance of the algorithms was measured by the t test. Results obtained from the survey pointed to the most productive state courts, such as the state of Rio de Janeiro, which has a large number of sentences, and the state of Pará, which has one of the fewest congestion numbers. The northeast region has the highest number of judgments by magistrates and the midwest has the least congestion. Despite the difficulty and lack of scientific methodology to measure the performance of the judiciary, the methodologies approached allow good comparisons and a comprehensive discussion.

**Keywords:** Data Mining, Judicial performance, state justice.



## LISTA DE ILUSTRAÇÕES

<b>Figura 1 - Processo KDD.....</b>	<b>15</b>
<b>Figura 2 - Sentenças proferidas por magistrados nas regiões brasileiras ...</b>	<b>18</b>
<b>Figura 3 - Produtividade do judiciário por região brasileira.....</b>	<b>19</b>
<b>Figura 4 - Sentenças proferidas por magistrados nos estados brasileiros ..</b>	<b>20</b>
<b>Figura 5 - Produtividade do judiciário por estado brasileiro.....</b>	<b>20</b>
<b>Figura 6 - Sentenças proferidas por magistrados entre de 2009 e 2017 .....</b>	<b>21</b>
<b>Figura 7 - Produtividade do judiciário por ano.....</b>	<b>22</b>
<b>Figura 8 - Quantidade de servidores.....</b>	<b>23</b>

## **LISTA DE TABELAS**

**Tabela 1 - Dimensões, categorias e variáveis revisadas nos estudos ....13**

**Tabela 2 - Teste  $t$  pareado dos algoritmos comparados com o PART.....24**

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>12</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>13</b>
<b>3</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>16</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>18</b>
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>24</b>
<b>6</b>	<b>REFERÊNCIAS .....</b>	<b>27</b>

## 1 INTRODUÇÃO

A avaliação de desempenho do judiciário, essencialmente, mensura a produção e produtividade dos tribunais e magistrados (GOMES; FREITAS, 2017). A atual demanda processual e o desejo por justiça e direitos da população impactam no número de casos novos no judiciário brasileiro, que comumente é associado a um setor lento e burocrático (GOIS; OLIVEIRA, 2017). Em 2004 a Emenda Constitucional nº 45, popularmente conhecida como Reforma do Judiciário, criou o CNJ, Conselho Nacional de Justiça (BRASIL, 2019).

O CNJ é “uma instituição pública que visa aperfeiçoar o trabalho do sistema judiciário brasileiro, principalmente no que diz respeito ao controle e à transparência administrativa e processual.” (CNJ, 2019a). O CNJ publica anualmente o relatório Justiça em Números que divulga “a realidade dos tribunais brasileiros, com muitos detalhes da estrutura e litigiosidade, além dos indicadores e das análises essenciais para subsidiar a Gestão Judiciária brasileira.” (CNJ, 2019b). Esses relatórios e dados são abertos ao público, sendo assim todo cidadão tem o direito de acessá-los livremente.

Com a disponibilização da base de dados da justiça brasileira, é possível aplicar algoritmos de mineração de dados para detectar padrões, anomalias, tendências e correlações nos dados dos tribunais de justiça dos 26 estados e do distrito federal. “Mineração de dados (em inglês, *data mining*) é o processo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados” (STATISTICAL ANALYSIS SYSTEM, 2019).

A preocupação com a gestão de recursos financeiros, a qualidade de serviços entregue pelo estado e o impacto social e humano causado pela ineficiência da justiça justificam e viabilizam esta pesquisa. Também é interessante conduzir essa pesquisa pois consegue-se trabalhar com dados reais e consistentes, para obter conclusões que ajudem na tomada de decisão dos administradores públicos. Nesse sentido, a inclusão de algumas técnicas de mineração de dados pode ampliar e contribuir com a avaliação de desempenho no judiciário.

Esta pesquisa tem como objetivo aplicar algoritmos de mineração dados nos determinantes do desempenho da justiça estadual brasileira nos dados disponíveis

entre os anos de 2009 e 2017. Para atingir esse objetivo, definiu-se os seguintes objetivos específicos: Identificar algoritmos para o desenvolvimento de modelos quantitativos; criar uma base de dados com as variáveis determinantes pela literatura e incluir outras disponíveis; avaliar o desempenho dos algoritmos; apresentar resultados consistentes do resultado da mineração de dados e propor e aplicar um método de avaliação do desempenho do judiciário.

## 2 REFERENCIAL TEÓRICO

A avaliação do desempenho do judiciário é um campo de estudo recente e que ainda deixa várias lacunas devido à um volume relativamente pequeno de produção científica, falta de concretização teórica e falta de metodologias para sistematizar o método de avaliação do desempenho, porém o assunto tem se difundindo (GOMES; GUIMARÃES, 2013; GOMES; FREITAS, 2017)

O desempenho do judiciário pode ser definido como “o processo sistemático de mensuração da performance de juízes, com objetivos institucionais notadamente voltados para o aprimoramento dos magistrados e do judiciário como um todo” (CNJ, 2011), assim é possível que os próprios magistrados consigam se avaliar e identificar seus pontos fracos e fortes, aperfeiçoando seu desempenho (CNJ, 2011).

Gomes e Guimarães (2013) revisaram os principais estudos de desempenho no judiciário e classificam as dimensões, categorias e variáveis de desempenho utilizadas na maior parte dos estudos. A quantidade de processos concluídos é apontada como a variável mais utilizada para avaliar a produtividade do judiciário. A Tabela 01 resume essas informações.

**Tabela 1 - Dimensões, categorias e variáveis revisadas nos estudos**

<b>Dimensão de desempenho</b>	<b>Principal categoria de interesse</b>	<b>Principais variáveis utilizadas</b>
<b>Eficiência</b>	Produtividade	<ul style="list-style-type: none"> <li>• Quantidade de processos concluídos</li> <li>• Quantidade de sentenças proferidas</li> </ul>
<b>Celeridade</b>	Duração de processos	<ul style="list-style-type: none"> <li>• Tempo de trâmite de procedimentos judiciais</li> <li>• Tempo de trâmite de procedimentos administrativos</li> </ul>
<b>Efetividade</b>	Confiança	<ul style="list-style-type: none"> <li>• Quantidade de violação aos direitos humanos</li> <li>• Quantidade de processos de corrupção</li> </ul>

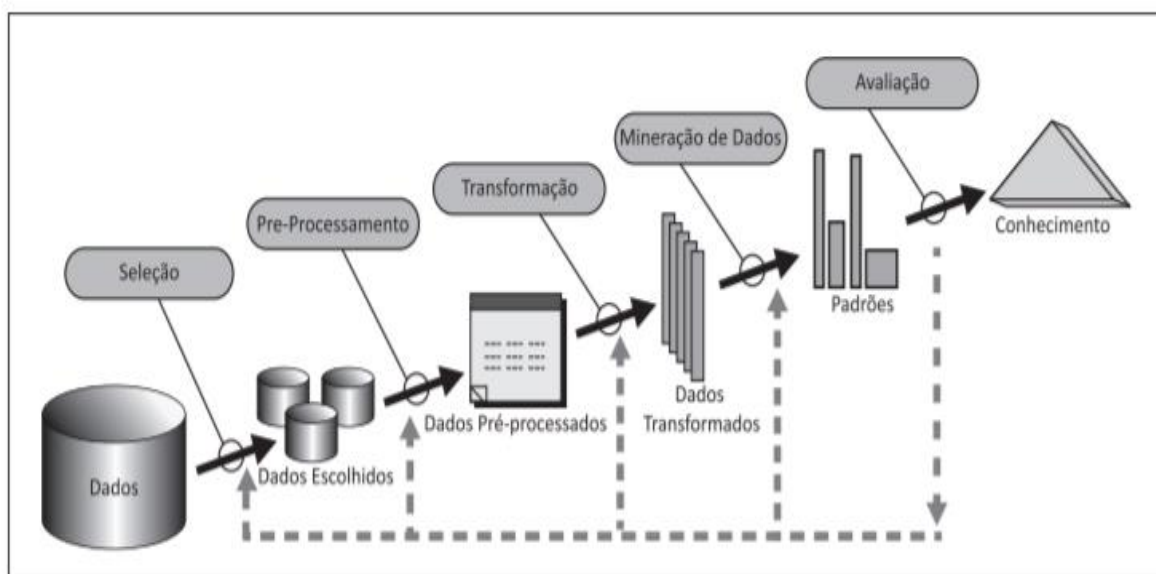
<b>Qualidade</b>	Mérito de decisões	<ul style="list-style-type: none"> <li>• Quantidade de decisões publicadas</li> <li>• Quantidade de decisões reformadas</li> </ul>
<b>Independência</b>	Autonomia	<ul style="list-style-type: none"> <li>• Quantidade de decisões contrárias ao governo</li> <li>• Quantidade de recursos financeiros alocados</li> </ul>
<b>Acesso</b>	Abrangência	<ul style="list-style-type: none"> <li>• Quantidade de juízes per capita</li> <li>• Quantidade de pessoas atendidas</li> </ul>

Fonte: Gomes e Guimarães (2013, p.645)

Gomes e Freitas (2017) analisaram por meio de estatística descritiva as seguintes variáveis: acervo total de processos, acervo recente de processos, quantidade de juízes, processos julgados por ano, audiências por ano, processos julgados por juiz (por ano) e audiências por juiz (por ano). Os resultados da análise dos dados de média, desvio padrão, mínimo e máximo mostraram valores discrepantes, desvio padrão alto e grandes diferenças nos dados das variáveis de mínimo e máximo. Outra técnica abordada pelos autores foi de verificar se há correlação entre as variáveis em estudo. Os resultados mais obtidos apontaram que: a) existe ausência de correlação entre a demanda de varas e quantidade de juízes; b) há correlação positiva entre a demanda e o desempenho das varas. c) os processos recentes apontam maior correlação com o desempenho do que o número total de processos. d) existe correlação negativa entre a quantidade de processos julgados e a quantidade de audiências efetuadas. e) há correlação negativa entre a quantidade de juízes e a produtividade das varas (GOMES, A. O.; FREITAS, 2017)

A mineração de dados é uma parte do processo de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases*) KDD, como é mais conhecido, é um processo de extração de informações uteis que consiste em 5 etapas: seleção de dados, pré-processamento, transformação, mineração e avaliação. O processo KDD permite a obtenção de conhecimento a partir dos passos deste processo que são apresentados na Figura 1. (FAYYAD et al; NAMEN, 2016)

**Figura 1 - Processo KDD**



**Fonte: FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1994**

Inicialmente é feito a seleção dos dados, esta seleção baseia-se em deixar apenas os dados que tem relevância e estão dentro das variáveis a serem estudadas na pesquisa. Logo após realiza-se a limpeza dos dados e o pré-processamento, que consiste em eliminar inconsistências dos dados e incongruências (caso existam) para que não influencie nos resultados. Logo após a etapa de transformação permite reduzir e adaptar os dados para um formato de melhor entendimento para o algoritmo que será usado na mineração de dados. Com os dados transformados é possível aplicar os algoritmos de mineração de dados e depois avaliar se os resultados estão dentro dos padrões definidos da mineração de dados. Caso não esteja deve-se voltar pra etapa de seleção e fazer novamente o processo de KDD (NAMEN, 2016).

As técnicas de mineração de dados permitem identificar padrões e tendências dado um conjunto de dados. Dentre as técnicas usadas na mineração de dados, uma das mais utilizada e comum é a de associação, que permite associar dados para, por exemplo, identificar tendências de compras. Outra técnica interessante também é a de classificação, com esta técnica é possível agrupar dados por classes, por exemplo: é possível agrupar clientes por idade e grupo social. A técnica de armazenamento em cluster é bastante similar à de classificação

e facilita o agrupamento de dados visualizar onde as semelhanças ou variações dos dados se encontram. A técnica de padrões sequenciais é mais utilizada em conjuntos de dados de longo prazo para identificar se há uma frequência regular de eventos similares (BROWN, 2012).

### 3 MATERIAIS E MÉTODOS

Esta seção descreve os materiais e métodos utilizados para conduzir a pesquisa, esta pesquisa pode ser classificada como abordagem quantitativa.

As variáveis usadas na pesquisa foram: ano, estado, região, PIB, casos novos, total de sentenças proferidas, despesas com comunicação e tecnologia da informação, quantidade de casos novos por magistrado, quantidade de sentenças proferidas por magistrado, quantidade de servidores e porte do tribunal.

A revisão das variáveis apresentadas na tabela 1 apontam a quantidade de sentenças proferidas como uma das variáveis mais utilizadas na literatura para mensurar a produtividade do judiciário. Essa variável pode ser trabalhada de várias formas, neste trabalho usa-se a quantidade de sentenças proferidas por magistrado, que é representada pela seguinte expressão:

$$\text{quantidade de sentenças proferidas por magistrado} = \frac{\text{quantidade de sentenças}}{\text{quantidade de juizes}} \quad 01$$

Se trabalhar apenas com a quantidade de sentenças proferidas despreza-se a força de trabalho dos magistrados, trabalhando desta forma é possível considerar o quão produtivo são os juizes.

Propõe se um segundo método de avaliar a produtividade do judiciário, dividindo a quantidade de casos novos por magistrado pela quantidade de sentenças proferidas por magistrado. Este modo de mensurar a produtividade não é apontado pela literatura, porém permite uma melhor comparação dos resultados e reforça a discussão pois se torna possível considerar o congestionamento dos tribunais, isto é, a quantidade de processos que ficam sem julgar.

A expressão matemática para mensurar a produtividade do judiciário será representada desta forma:



$$\text{produtividade do judiciário} = \frac{\text{quantidade de casos novos por magistrado}}{\text{quantidade de sentenças proferidas por magistrado}} \quad 02$$

É necessário considerar que o resultado dessa expressão indicará um valor que, quando mais próximo de 0, apresentará um melhor resultado da produtividade do judiciário, conseqüentemente um valor distante de 0 indica uma produtividade menor. Também é interessante considerar que quando obtemos um valor menor que 1 significa que a quantidade de casos novos por magistrado é menor que a quantidade de sentenças proferidas por magistrado. Quando tivermos um valor igual a 1 a quantidade de casos novos por magistrado é exatamente igual a quantidade de sentenças proferidas por magistrado, e por sua vez quando o valor for maior que 1 a quantidade de casos novos por magistrado é maior que a quantidade de sentenças proferidas por magistrado.

Os dados são disponibilizados em formato CSV via *download* pelo portal do CNJ, utilizou-se o editor de planilhas eletrônicas Libre Office Calc para o pré-processamento de dados e a aplicação dos algoritmos de mineração foram realizadas por meio do software Weka versão 3.9. A classificação dos dados foi feita por meio regra de Sturges, sugerida por PACITTI E ATKINSON (1997), que é representada pela fórmula  $k = 1 + 3,322 * \log(n)$  onde  $k$  representa o número de classes e  $n$  a quantidade de amostras. Tem-se 243 registros para cada variável. Esse valor representa o total amostras que deve ser representado por  $n$ . Assim temos:

$$k = 1 + 3,322 * \log(n)$$

$$k = 1 + 3,322 * \log(243)$$

$$k = 1 + 3,322 * 2,386$$

$$k \cong 8,925$$

Como obteve-se um valor próximo a 9, a classificação distribui os dados em 9 faixas de dados. O tratamento de distribuição de frequências que substitui o valor numéricos em classes de dados foi realizado no Sistema Gerenciador de Banco de Dados (SGBD) MySQL em sua versão 5.7.

Os resultados obtidos por meio do processo de extração de conhecimento foram apresentados em tabelas e gráficos. Para extrair esse conhecimento foram utilizadas as técnicas de mineração de dados de classificação e associação.

A técnica de classificação foi feita com os algoritmos PART, OneR e J48, para a associação será usado o algoritmo Apriori. Como a base de dados do CNJ é sólida e consistente não se encontra problemas de *missing values* (valores faltantes).

Para fins de comparação do desempenho do judiciário é interessante comparar os resultados dos tribunais agrupados por estado (comparar os tribunais estaduais entre si) e por região (norte, sul, centro-oeste, sudeste e nordeste) e por ano (de 2009 a 2017) para buscar entender as diferenças entre os tribunais.

A avaliação da performance dos algoritmos foi feita pelo teste t realizado no próprio Weka, que deverá indicar qual algoritmo é melhor estatisticamente em comparação ao outro. O nível de significância especificado para o teste foi de 0,05.

#### 4 RESULTADOS E DISCUSSÕES

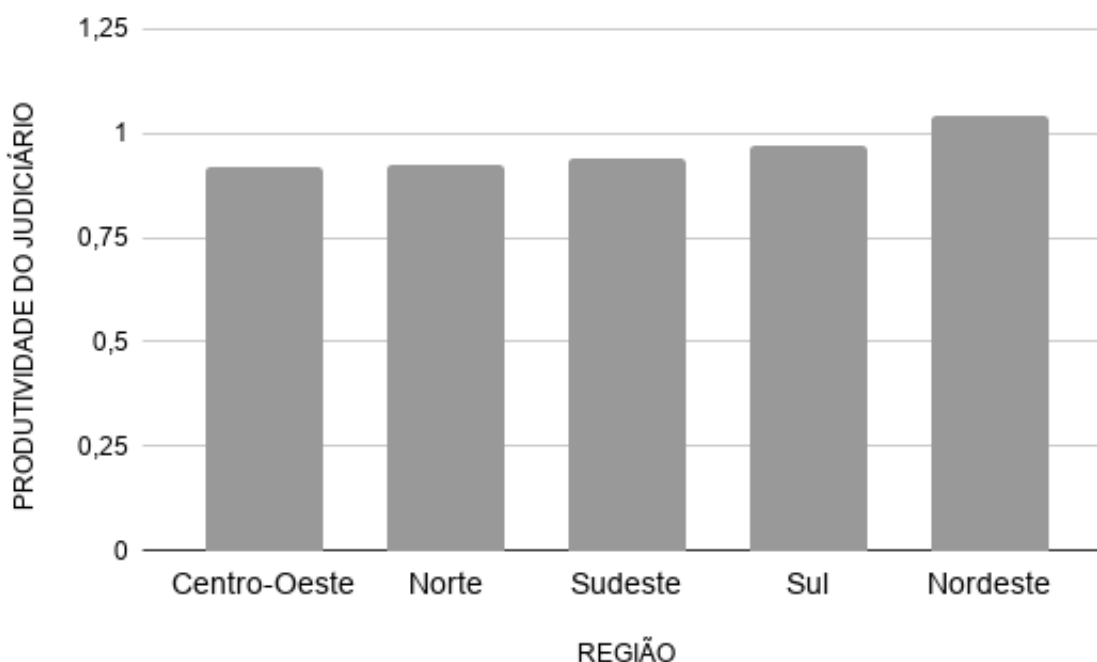
Nesta seção apresenta-se os resultados obtidos por meio dos cálculos realizados. Os resultados serão descritos em duas partes: do processo de extração de conhecimento, tratando da parte que mensura o desempenho da justiça estadual e do processo de avaliação dos algoritmos usados para realização dos experimentos.

**Figura 2 - Sentenças proferidas por magistrados nas regiões brasileiras**



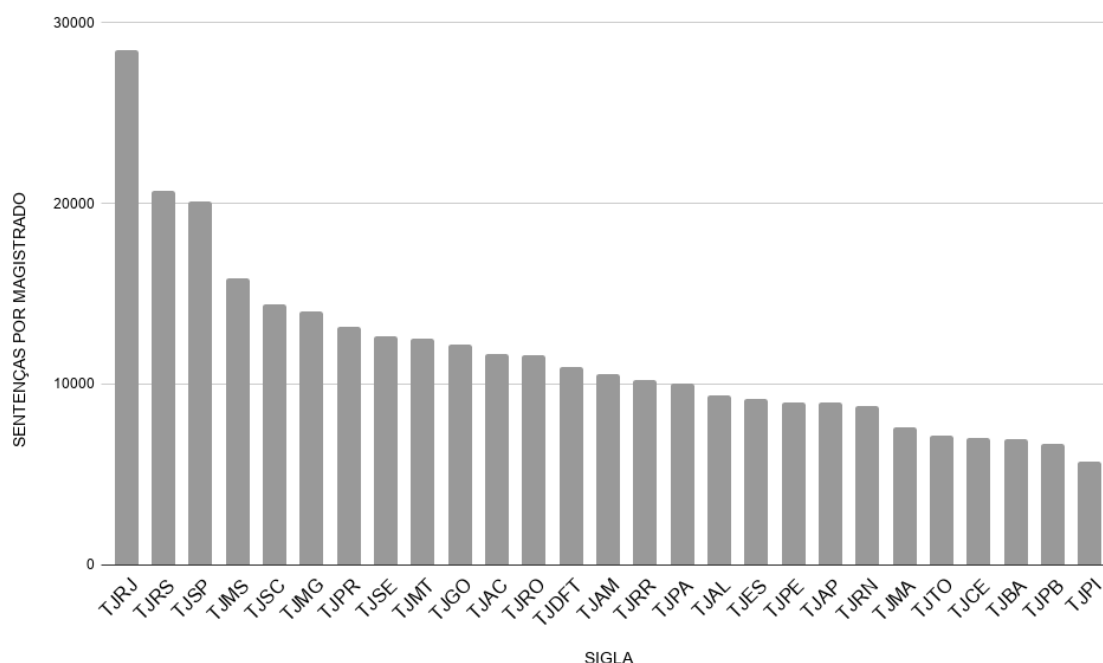
A Figura 2 demonstra a quantidade de sentenças total proferidas por magistrados nas regiões brasileiras entre os anos de 2009 e 2017, somando todas as sentenças desse intervalo de tempo de todos os tribunais que compõem as regiões brasileiras. Porém pode ser equivocado mensurar o desempenho dos tribunais considerando apenas as sentenças proferidas por cada magistrado e desprezar os casos novos para cada magistrado, assim ignorando a taxa congestionamento criada. Abaixo, A Figura 2 apresenta os casos novos e assim não despreza o congestionamento criado nos tribunais.

**Figura 3 - Produtividade do judiciário por região brasileira**



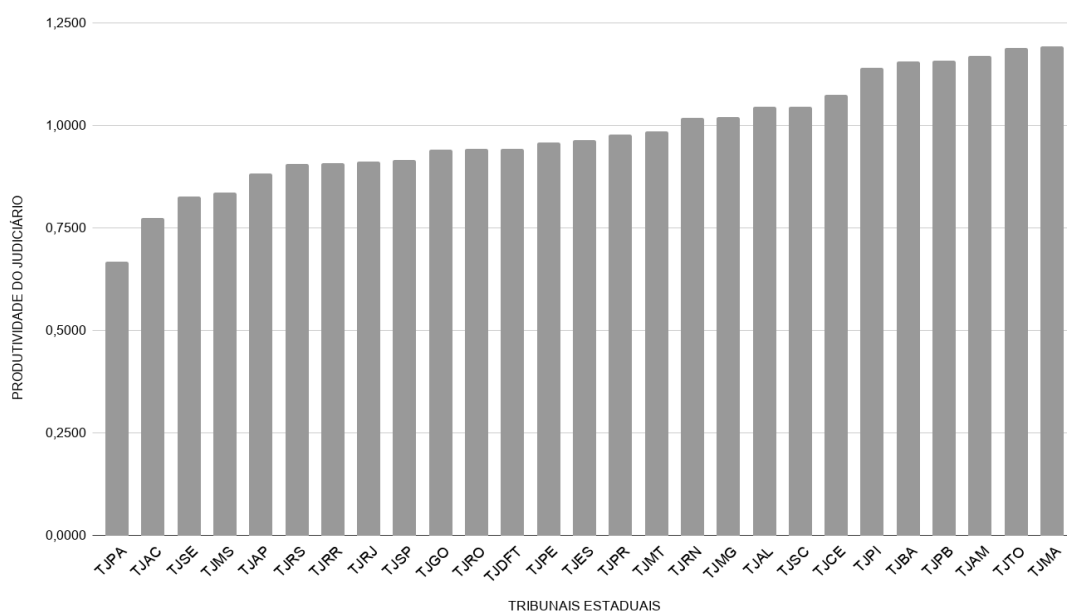
Na Figura 3 apresenta-se a produtividade do judiciário agrupado por cada região do Brasil entre os anos de 2009 e 2017, neste gráfico apresenta-se os resultados considerando os novos casos por magistrado, vale ressaltar que um valor mais próximo a 0 indica um melhor desempenho. Desta forma existe uma mudança no cenário de qual região é mais eficiente. Se tomar como exemplo a região Nordeste, no primeiro caso quando se despreza os casos novos por magistrado temos o Nordeste como a região mais eficiente, porém quando consideramos a quantidade de casos novos para cada magistrado o cenário muda radicalmente e a região Nordeste mostra-se a menos eficiente de todas. A seguir agrupamos por estado:

**Figura 4 - Sentenças proferidas por magistrados nos estados brasileiros**



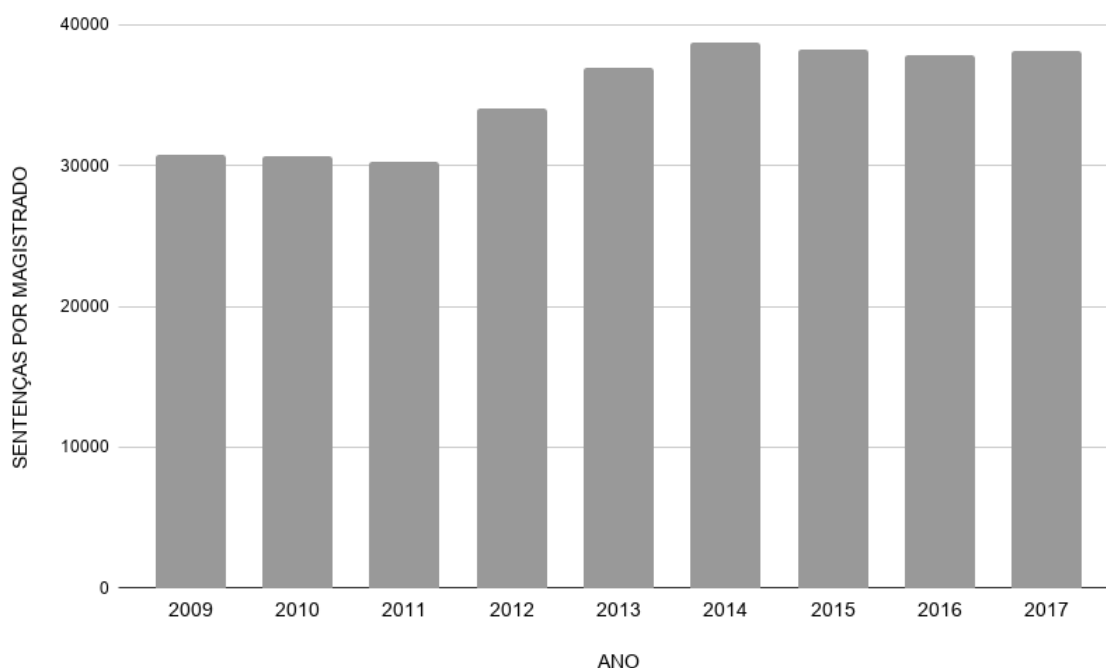
A figura 4 demonstra a quantidade de sentenças proferidas pelos magistrados nos estados brasileiros entre os anos de 2009 e 2017, somando todas as sentenças deste intervalo de tempo. O TJRJ (Tribunal de Justiça do Rio de Janeiro) chama atenção pelo grande número de sentenças proferidas por magistrados, o colocando como o melhor quando comparado aos outros estados. Abaixo, na figura 5, trata-se da produtividade do judiciário considerando os casos novos.

**Figura 5 - Produtividade do judiciário por estado brasileiro**



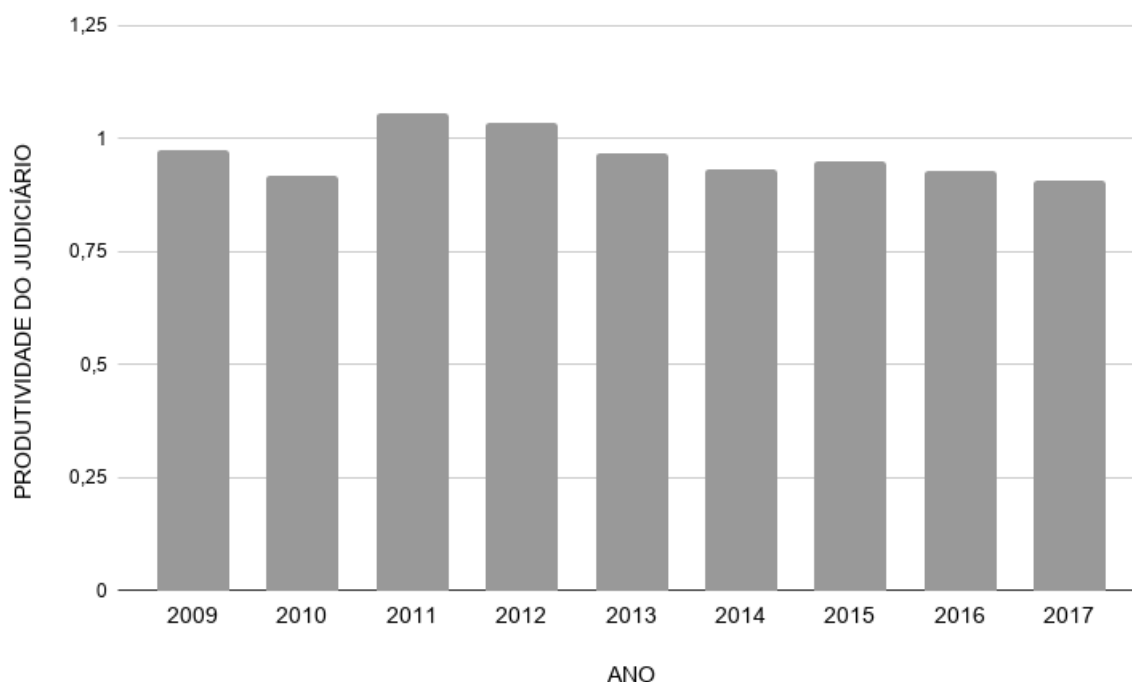
A Figura 5 apresenta a mesma ideia da Figura 3, porém agrupando-se por estados ao invés de regiões, sendo assim os estados com a produtividade do judiciário mais próximos a 0 representam um melhor desempenho, na apresentação gráfica temos da esquerda para a direita representando como do melhor para o pior. Dois exemplos interessantes a discutir são: o TJPA (Tribunal de Justiça do estado do Pará) se mostra muito eficiente quando se considera os casos novos, se colocando em primeiro lugar, porém quando é desprezado o número de casos novos do TJPA e conta-se apenas as sentenças proferidas por magistrado, o tribunal sai do primeiro lugar e ocupa a 16ª posição. O outro é o TJRJ que apresenta um grande número de sentenças por magistrado, porém quando se considera os casos novos, o TJRJ sai do primeiro lugar e vai para a 8ª posição. Abaixo, na figura 6 agrupamos por ano.

**Figura 6 - Sentenças proferidas por magistrados entre de 2009 e 2017**



Na Figura 6 é representando a quantidade de sentenças proferidas entre os anos de 2009 e 2017, contando todos os tribunais estaduais brasileiros. Apesar de o ano de 2014 ser o mais produtivo dentre os anos analisados, é possível perceber uma diferença pequena principalmente entre os anos. Neste caso não está sendo considerado a quantidade de casos novos, assim desprezando o congestionamento criado nesses anos. A seguir, na Figura 7, considera-se os casos novos.

**Figura 7 - Produtividade do judiciário por ano**

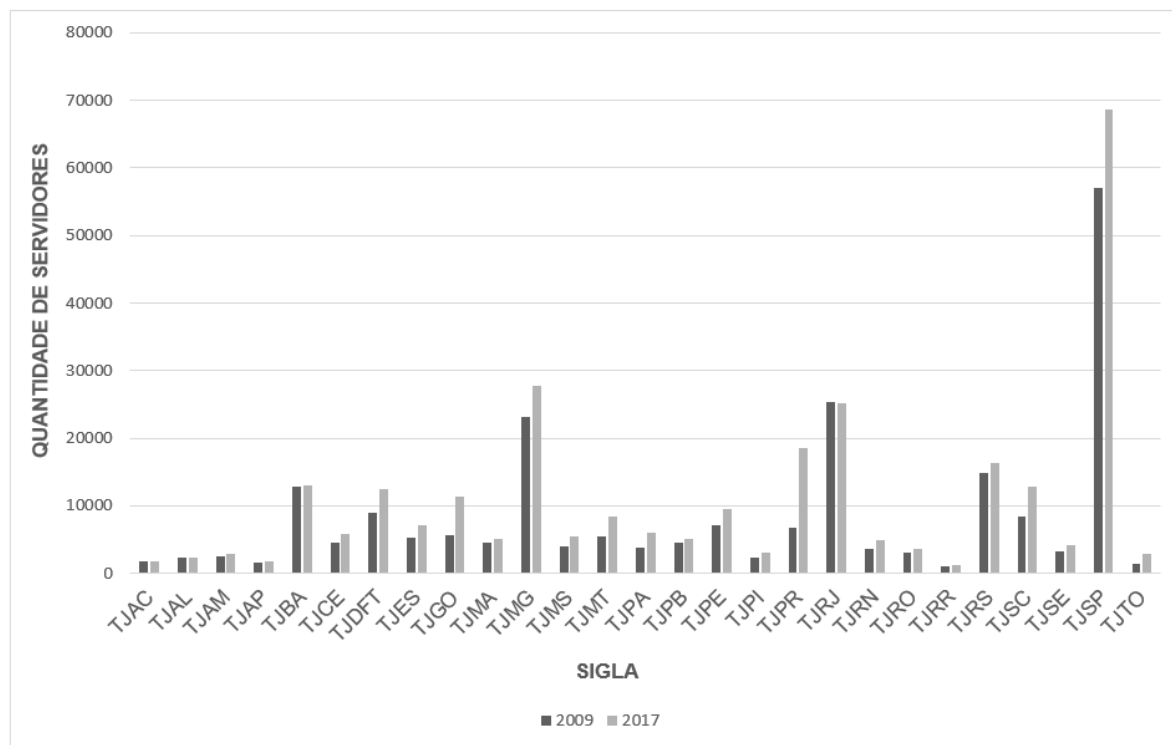


Na Figura 7 o gráfico apresenta a produtividade do judiciário agrupada por ano, todos os anos com exceção de 2012 e 2011 obtiveram um resultado menor do que 1, que indicam que tem um número de sentenças maior do que de casos novos. Os anos de 2012 e 2011 como apresentam um resultado maior que 1 indicam que houve mais casos novos do que sentenças proferidas por magistrado. O ano que mais chama atenção é o de 2010, que quando não se considera os casos novos aparece em penúltimo lugar, porém quando considera-se os casos novos já aparece como o segundo ano mais produtivo.

Com a aplicação dos algoritmos de mineração de dados foi possível identificar padrões, indicando que: a) Os tribunais de grande porte têm a melhor produtividade (1,0755), seguido dos tribunais pequenos (1,0379), e por fim, os tribunais de médio porte indicam o pior desempenho (1,0755). b) Os tribunais de pequeno porte são os que tem o maior número de sentenças proferidas por magistrados (119814 sentenças), seguido dos de médio porte (104070 sentenças) e por fim os tribunais de grande porte (91602 sentenças). c) O algoritmo Apriori encontrou regras de associação que indicam que os tribunais que mais investem em comunicação Tecnologia da Informação tendem a ser mais produtivos. d) Não se encontrou

associações que indicam que os tribunais dos estados que têm o maior PIB sejam mais produtivos.

**Figura 8 - Quantidade de servidores**



A Figura 8 apresenta um comparativo da quantidade total de servidores dos tribunais estaduais nos anos 2009 e 2017, que é o primeiro e o último ano avaliado no estudo. Os resultados apontam que a quantidade de servidores aumentou na maioria dos tribunais, entre eles, os tribunais que mais se destacam quanto ao aumento do número de servidores são: TJSP (Tribunal de Justiça do estado de São Paulo), TJPR (Tribunal de Justiça do estado do Paraná), TJDFT (Tribunal de Justiça do Distrito Federal e dos Territórios), TJGO (Tribunal de Justiça do estado de Goiás), TJSC (Tribunal de Justiça do estado de Santa Catarina), TJMG (Tribunal de Justiça do estado de Minas Gerais) TJPE (Tribunal de Justiça de Pernambuco). TJSP, TJRJ, TJMG são 3 tribunais que têm o maior número de servidores.

O software Weka permite a avaliação do desempenho dos algoritmos de classificação por meio do teste *t*. Na tabela abaixo observamos o desempenho de mais nove algoritmos quando comparados com o PART:

**Tabela 2 - Teste *t* pareado dos algoritmos comparados com o PART**

<b>Algoritmo</b>	<b>Desempenho</b>	<b>Resultado estatístico (weka)</b>	<b>Resultado estatístico</b>
PART	62,69%	v	Melhor
One R	61,05%		Não há diferença
REPTRee	61,05%		Não há diferença
J48	60,76%		Não há diferença
Randon Tree	59,66%		Não há diferença
NaiveBayes	57,71%	*	Pior
NaiveBayesUpdateable	57,71%	*	Pior
DecisionTable	57,14%	*	Pior
Zero R	38,24%	*	Pior
NaiveBayesMultinomialText	38,24%	*	Pior

Fonte: Dados da pesquisa

Como apresentado na Tabela 2, o algoritmo PART tem um desempenho melhor que o OneR e o J48 (algoritmos usados no experimento para extração de conhecimento do desempenho do judiciário). Se todos os algoritmos apresentados tivessem sido usados para obter resultados da produtividade do judiciário, eles ficariam ordenado como na tabela, do melhor para o pior. O Weka utiliza v para indicar o melhor resultado e \* para indicar o pior resultados, quando esses valores (v ou \*) não aparecem significa que não existe diferenças estatística.

O algoritmo PART é estatisticamente superior a todos os outros, porém não existe diferença significativa estatística quando comparados o OneR e o J48. Os algoritmos REPTRee e RandonTree também não apresentam diferença significativa quando comparado ao PART. NaiveBayes, NaiveBayesUpdateable, DecisionTable, Zero R, NaiveBayesMultinomialText tem o pior resultado estatístico quando comparados ao PART, assim se mostram menos eficientes e se tornam os piores em questão de desempenho.

Não é possível aplicar o mesmo parâmetro de comparação para o algoritmo de associação Apriori que também foi utilizado para o tratamento de dados nesta pesquisa.

## **5 CONCLUSÃO**

Como foi mencionado ao longo deste trabalho, o conceito de avaliação do desempenho do judiciário é relativamente novo e pouco explorando ainda, falta uma



metodologia sistemática de como realmente o desempenho do judiciário deve ser avaliado, quais parâmetros devem ser utilizados, e quais variáveis devem ser mais ou menos consideradas, deixando-se assim lacunas na literatura sobre um modelo sistêmico de avaliação da produtividade do judiciário, isto é, o modo de se avaliar o judiciário. Também não há critérios para ponderar a importância das variáveis e atribuir pesos a elas.

O método de avaliar o desempenho do judiciário por meio da quantidade de sentenças proferidas por juízes permite boas comparações entre os tribunais, permitindo comparar diversos tipos de grupos, a depender da criatividade do pesquisador e do conjunto de dados disponível. O método proposto no trabalho de avaliar os casos novos dividido pela quantidade de sentenças se limita a comparar de diferentes perspectivas o desempenho do judiciário e ampliar a discussão de diferentes óticas, porém como não existe metodologias consolidadas para avaliação do desempenho do judiciário, fica-se refém desses métodos. Por outro lado, as metodologias de mineração de dados já são mais exploradas e difundidas, o que nos permite garantir resultados consistentes independente do conjunto de dados tratado. O processo KDD também já é bastante conhecido e utilizado para extração de conhecimento.

O que muito se existe na literatura são estudos que propõe entender por qual motivo o judiciário é mais ou menos produtivo, por exemplo, Gomes e Guimarães (2013) buscam saber se juízes com maior idade são mais produtivos ou não do que os juízes com idades inferiores, podendo justificar pela inserção de novas tecnologia da informação e sua dificuldade de lidar com tecnologias. Outro exemplo é avaliar se a inserção de mais juízes melhora o desempenho, uma vez que pode ocorrer de piorar o desempenho ao diminuir a carga de trabalho e pressão sobre os juízes antigos e causando uma possível redução no desempenho, porém é difícil obter dados pessoais e individuais dos magistrados como idade, instrução e afinidade com novas tecnologias, por não conter na base de dados disponibilizada pelo CNJ, assim inviabilizando esses tipos estudos.

As limitações da pesquisa não causaram impacto nos resultados, ainda com a limitação da falta de metodologia para avaliação do desempenho do judiciário, foi possível discutir resultados e comparar a produtividade do judiciário por região,

estado e ano. A limitação de mensurar o desempenho do Apriori também não atrapalhou nos resultados, apenas os excluiu do comparativo.

Aproveita-se a oportunidade para propor ao CNJ que inclua em suas bases, dados individuais sobre os magistrados como especialidades em área de conhecimento, idade e tempo de carreira. A inserção do nível de escolaridade dos membros envolvidos no processo também permite trabalhos interessantes com mineração de dados, possibilitando estudar se pessoas com mais escolaridade buscam mais ou menos por justiça. Aos pesquisadores das áreas de direito e administração pública recomenda-se que definam um parâmetro para avaliação da produtividade do judiciário.

Contudo, os objetivos da pesquisa foram todos alcançados, aplicou-se algoritmos de mineração de dados nos determinantes do desempenho da justiça entre os anos de 2009 e 2017. Construiu-se uma base de dados com as variáveis apontadas pela literatura também incluindo de outras disponíveis, apresentou-se os resultados do processo de mineração de dados, avaliou-se o desempenho dos algoritmos, propôs-se um método de avaliação e atribui-se os dados aos parâmetros propostos.

## 6 REFERÊNCIAS

BROWN, M. *IBM*. Disponível em: <<https://www.ibm.com/developerworks/br/library/tecnicas-mineracao-de-dados/index.html>>. Acesso em: 22 abr. 2019.

CONSELHO NACIONAL DE JUSTIÇA. *Avaliação do desempenho judicial: desafios, experiências internacionais e perspectivas*. Disponível em: <<http://www.cnj.jus.br/images/pesquisas-judiciarias/Publicacoes/40-211-1-PB.pdf>>. Acesso em: 15 abr. 2019.

CONSELHO NACIONAL DE JUSTIÇA. Disponível em: <<http://www.cnj.jus.br/sobre-o-cnj/quem-somos-visitas-e-contatos>>. Acesso em: 25 mar. 2019a.

CONSELHO NACIONAL DE JUSTIÇA. Disponível em: <<http://www.cnj.jus.br/pesquisas-judiciarias/justicaemnumeros/2016-10-21-13-13-04/pj-justica-em-numeros>>. Acesso em: 25 mar. 2019b.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Attempt to isolate mast-cell precursors based on the differential sensitivity to UV-B and X-irradiation. *Toxic Substances Journal*, v. 13, n. 2, p. 85–95, 1994.

GOIS, L.; OLIVEIRA, L. Dez anos de CNJ: Reflexões do envolvimento com a melhoria da eficiência do Judiciário brasileiro. *Rev. Serv. Público Brasília*, v. 68, n. 3, p. 631–656, 2017.

GOMES, A. DE O.; GUIMARÃES, T. DE A. Desempenho no Judiciário: conceituação, estado da arte e agenda de pesquisa. *Revista de Administração Pública*, v. 47, n. 2, p. 379–401, 2013.

GOMES, A. O.; FREITAS, M. E. M. Correlação entre demanda, quantidade de juízes e desempenho judicial em varas da Justiça Federal no Brasil. *Revista Direito GV*, v. 13, n. 2, p. 567–585, 2017.

NAMEN, A. A. MINERAÇÃO EM BASES DE DADOS DO INEP: UMA ANÁLISE EXPLORATÓRIA PARA NORTEAR MELHORIAS NO O século XXI se caracteriza por um cenário em que as atividades humanas se organizam em torno da geração, recuperação e utilização do conhecimento. Esse novo contex. p. 133–157, 2016.

PACITTI, T.; ATKINSON, C. P. *Programação e métodos computacionais*. 2. ed. Rio de Janeiro - RJ: Livros Técnicos e Científicos Editora S/A, 1997.

*PLANALTO.* Disponível em:  
<[http://www.planalto.gov.br/ccivil\\_03/constituicao/emendas/emc/emc45.htm](http://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc45.htm)>. Acesso em: 9  
abr. 2019.

*STATISTICAL ANALYSIS SYSTEM.* Disponível em:  
<[https://www.sas.com/pt\\_br/insights/analytics/mineracao-de-dados.html](https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html)>. Acesso em: 25  
mar. 2019.