

## **OUVIDORIA-CIDADÃ: Uma Proposta Metodológica Baseada Em Inteligência Artificial**

**CAROLINA COELHO DA SILVEIRA**  
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

**CARLA BONATO MARCOLIN**  
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

**GABRIEL AUGUSTO SAMPAIO SOARES**  
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

**GABRIELA LOTTA**  
FUNDAÇÃO GETULIO VARGAS/EBAPE

**EDUARDO DE REZENDE FRANCISCO**  
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

### **Agradecimento à órgão de fomento:**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

# **OUVIDORIA-CIDADÃ: Uma Proposta Metodológica Baseada Em Inteligência Artificial**

## **1. INTRODUÇÃO**

A Administração Pública contemporânea vive um momento de transformação digital sem precedentes, caracterizado pela crescente integração de tecnologias emergentes aos processos de gestão pública e pela migração progressiva dos serviços presenciais para as plataformas digitais (Mergel *et al.*, 2023). O uso de Inteligência Artificial (IA) tem ganhado destaque como uma das principais tendências dessa transformação, oferecendo ferramentas capazes de otimizar processos, personalizar o atendimento aos cidadãos e aprimorar a tomada de decisões com base na extração de informações de grandes volumes de dados públicos (Madupati, 2024).

Essa transição vai além de uma mudança técnica: está reconfigurando a relação entre Estado e sociedade e alterando, além dos mecanismos de prestação de serviços, também canais de controle e de participação social (Breit *et al.*, 2020; Lindgren *et al.*, 2019). Com essa materialização dos encontros públicos no ambiente digital, passa a existir a necessidade de processar e analisar eficientemente os crescentes volumes de dados gerados por essas novas formas de interação.

Particularmente, os dados textuais adquirem cada vez mais relevância visto que fornecem informações sobre as necessidades, percepções e demandas da população, exigindo abordagens analíticas mais sofisticadas que possam extrair conhecimento útil desse conjunto de dados não estruturados (Breit *et al.*, 2020; Lindgren *et al.*, 2019; Madupati, 2024). Neste cenário, as técnicas Processamento de Linguagem Natural (NLP) surgem como alternativa promissora, possibilitando a identificação automatizada de padrões, tendências e *insights* que talvez passassem despercebidos em análises tradicionais.

Diante deste contexto, o presente estudo tem como objetivo apresentar uma proposta metodológica baseada em técnicas de Processamento de Linguagem Natural para análise automatizada das manifestações encaminhadas a ouvidorias públicas, com validação empírica utilizando dados das ouvidorias conectadas à Ouvidoria Geral da União do Brasil. A partir do método proposto, espera-se oferecer uma alternativa inovadora para a extração de informações sobre as interações entre cidadãos e Estado, contribuindo para uma melhor compreensão dos desafios e oportunidades presentes na prestação de serviços públicos.

## **2. CONTEXTO INVESTIGADO E DIAGNÓSTICO DA SITUAÇÃO-PROBLEMA**

Organizações do setor público em diferentes países estão adotando tecnologias digitais para apoiar e, em muitos casos, substituir os encontros presenciais entre cidadãos e funcionários públicos. Essa digitalização é impulsionada por duas forças principais: a busca pelo autoatendimento, permitindo que cidadãos acessem dados e serviços governamentais de forma mais conveniente, e a necessidade de eficiência interna através da automação de processos administrativos (Lindgren *et al.*, 2019).

Estudos recentes (Eckhard; Friedrich, 2024; Gilad *et al.*, 2024) revelam que a maior parte das interações entre cidadão e Estado atualmente são, de fato, digitais, exigindo a extensão da abordagem relacional de estudo do encontro público para o domínio das comunicações virtuais. Nesse contexto, as ouvidorias públicas emergem como instrumentos de mediação entre governo e cidadãos.

No Brasil, a institucionalização dessas estruturas organizacionais culminou com a criação da Ouvidoria Geral da União (OGU), que centraliza e coordena as atividades de ouvidoria em âmbito federal (Controladoria Geral da União, 2019). Definidas como canais formais que permitem aos cidadãos manifestarem demandas, reclamações, sugestões e denúncias sobre serviços públicos (Controladoria Geral da União, 2019), as ouvidorias

funcionam como uma ponte entre Estado e sociedade e são responsáveis por processar volumes exponenciais de manifestações através de plataformas digitais integradas.

A digitalização das ouvidorias, embora ofereça novas oportunidades para o atendimento, apresenta também desafios significativos para a análise sistemática das manifestações cidadãs. Por um lado, a facilidade de acesso e conseqüente crescimento no volume de manifestações recebidas representa uma rica fonte de informações sobre percepções, necessidades e experiências cidadãs em relação aos serviços públicos, mas demanda novas abordagens metodológicas capazes de processar e analisar sistematicamente esses grandes volumes de dados textuais não estruturados.

Não somente o volume de manifestações, mas a diversidade temática das manifestações, que abrangem desde questões burocráticas específicas até denúncias sobre políticas públicas, exige abordagens analíticas capazes de capturar essa heterogeneidade. Simultaneamente, a qualidade dos dados textuais (incluindo manifestações duplicadas, incompletas ou inadequadas) compromete a efetividade das análises quando não tratada adequadamente. Por fim, a necessidade de responsividade das ouvidorias, que precisam fornecer respostas ágeis e adequadas às manifestações, demanda ferramentas analíticas capazes de identificar padrões e priorizar atendimentos.

Portanto, essa é uma tarefa que se torna muito desafiadora através de métodos tradicionais de análise de dados de texto. Esses métodos baseiam-se, predominantemente, em estatísticas descritivas simples e regras linguísticas pré-definidas, como contagem de frequências de palavras, análise de co-ocorrências e processamento baseado em gramáticas formais (Postiglione, 2024). Por se basearem em estruturas rígidas da linguística, podem não capturar adequadamente as nuances semânticas e contextuais da linguagem natural, seja ela falada ou escrita.

Por conseguinte, essa abordagem apresenta limitações significativas, como: a subjetividade nas análises sujeitas a interpretações pessoais que podem gerar inconsistências entre diferentes analistas; a dificuldade do processamento em escala, já que o volume de manifestações cresce exponencialmente à medida que as organizações expandem sua presença digital e os canais de comunicação com o público se multiplicam e se tornam mais acessíveis; e classificações pré-definidas, que podem não capturar a complexidade das demandas cidadãs.

Por isso, novas oportunidades tecnológicas associadas à mineração de dados e aprendizado de máquina têm sido discutidas com crescente interesse por acadêmicos e gestores públicos (Geske; Leyer, 2022). Essas tecnologias emergentes podem potencialmente cumprir objetivos centrais do governo digital, incluindo melhoria da eficiência e qualidade dos serviços, redução dos tempos de espera, aumento da transparência e oferecimento de prestação de serviços integrada entre organizações (Alshahrani *et al.*, 2022).

Diante desse contexto de grande volume de dados não-estruturados gerados pelas ouvidorias digitais, a Inteligência Artificial emerge como uma opção promissora para o processamento e análise sistemática dessas informações. Essas abordagens computacionais não substituem, mas complementam e expandem significativamente as possibilidades analíticas das metodologias tradicionais das ciências sociais, oferecendo novas perspectivas para a compreensão das dinâmicas de interação entre Estado e sociedade. Além disso, a aplicação dessas tecnologias no contexto das ouvidorias públicas representa uma oportunidade para desenvolver sistemas mais responsivos e efetivos de prestação de serviços públicos, permitindo respostas mais ágeis às demandas cidadãs e identificação proativa de problemas sistemáticos nos serviços governamentais.

Contudo, é fundamental considerar as implicações éticas e democráticas dessas transformações tecnológicas. A implementação de sistemas de análise automatizada deve assegurar que a automação e a análise algorítmica não comprometam princípios fundamentais como transparência, *accountability* e participação que sustentam a administração pública

democrática (Cholyshkina *et al.*, 2024). A literatura também identifica desafios inerentes ao exercício da discricionariedade automatizada, particularmente a possibilidade de que algoritmos reproduzam vieses inconscientes presentes nos dados de treinamento, levando a tratamentos desiguais baseados em características como origem geográfica, tipo de linguagem utilizada ou complexidade das demandas (Lindgren *et al.*, 2019; Peeters, 2023)

Diante desta situação, identifica-se a oportunidade de desenvolver metodologias que combinem o potencial analítico das tecnologias de Inteligência Artificial com os princípios democráticos que devem orientar a Administração Pública. A proposta metodológica apresentada neste estudo busca contribuir para a aplicação responsável e transparente dessas tecnologias no contexto das ouvidorias públicas brasileiras, oferecendo uma alternativa escalável e replicável para análise sistemática de manifestações cidadãs.

### 3. INTERVENÇÃO PROPOSTA

A metodologia proposta baseia-se em técnicas de Processamento de Linguagem Natural (NLP, do inglês *Natural Language Processing*), área de Inteligência Artificial cujo objetivo é realizar o processamento de grandes volumes de texto de maneira semelhante a um humano, buscando extrair resultados utilizáveis (Chatterjee, 2017). Ou seja, sua tarefa é semelhante ao papel de um leitor ou ouvinte. Recentemente, esse conjunto de técnicas tem se destacado pelo desenvolvimento de ferramentas de Inteligência Artificial Generativa, como o ChatGPT e o DeepSeek, e pela sua crescente flexibilização para se adaptar e interpretar inconsistências decorrentes de erros de pronúncia, sotaques, jogos de palavras, dialetos, contexto e outros (Banh; Strobel, 2023).

Os dados utilizados neste estudo foram fornecidos pela Ouvidoria Geral da União (OGU), já devidamente anonimizados e organizados em planilhas. O conjunto de dados compreende manifestações cidadãs encaminhadas a quatro ministérios: Ministério do Desenvolvimento Agrário e Agricultura Familiar (MDA), Ministério da Cultura (MinC), Ministério do Meio Ambiente e Mudança do Clima (MMA) e Ministério da Pesca e Aquicultura (MPA). Esta amostra foi selecionada por representar diferentes áreas temáticas da Administração Pública Federal, permitindo avaliar a aplicabilidade da metodologia proposta em contextos diversos de políticas públicas.

Originalmente, a base contava com 4.326 manifestações, divididas entre os quatro ministérios, conforme a Tabela 1 abaixo.

**Tabela 1 – Quantidade de Manifestações por Ministério**

Ministério do Desenvolvimento Agrário e Agricultura Familiar	673
Ministério da Cultura	1.400
Ministério do Meio Ambiente e Mudança do Clima	802
Ministério da Pesca e Aquicultura	1.451
<b>TOTAL</b>	<b>4.326</b>

Dados de texto, em geral, são desestruturados, amorfos e volumosos, exigindo preparação antes da aplicação das técnicas. Esse processo envolvendo processar os dados, removendo qualquer ocorrência que possa causar ruídos e convertendo esses dados para um *corpus* de formato aceitável para os algoritmos (Marcolin *et al.*, 2023; Sarkar, 2019).

Para este caso, optou-se por utilizar a linguagem computacional Python, amplamente utilizada para análise de dados em massa. Além disso, é uma linguagem que consegue lidar excepcionalmente bem com dados em texto, o que levou ao desenvolvimento de diversas bibliotecas e, conseqüentemente, diversas possibilidades de análise (Sarkar, 2019).

A fase de processamento deste projeto incluiu as melhores práticas indicadas pela literatura (Sarkar, 2019), como redução do léxico, remoção de *stopwords* e tokenização. Já para

as análises, foram utilizadas técnicas exploratórias de análise de janelas de palavras (*n-grams*), análise sintática (*PoS*, do inglês *Part-of-Speech tagging*) e uma análise não-supervisionada, a modelagem de tópicos (*topic modeling*). A combinação dessas técnicas permitiu uma abordagem mais completa considerando a natureza dos dados e o ineditismo da iniciativa.

A análise de *n-grams* auxilia a compreensão do contexto e frequência das principais expressões empregadas pelos cidadãos. Essa abordagem identifica sequências de *n* termos consecutivos no texto, permitindo capturar padrões e combinações de palavras recorrentes. Por exemplo, bigramas ( $n=2$ ) e trigramas ( $n=3$ ) ajudam a destacar frases ou termos compostos frequentemente mencionados, contribuindo para uma exploração mais precisa e contextualizada das demandas e preocupações presentes nos relatos analisados.

A técnica de *PoS* é aplicada para identificar e classificar as palavras de acordo com suas categorias gramaticais, como substantivos, verbos, adjetivos e advérbios. Essa abordagem permite compreender a estrutura das frases e a relação entre as palavras, destacando quais termos são frequentemente associados a problemas específicos (substantivos) ou a ações demandadas (verbos), bem como combinações de pares substantivos-verbos, substantivos-adjetivos, entre outros. Ao analisar as classes gramaticais, é possível mapear padrões linguísticos que refletem as principais solicitações, enriquecendo a interpretação dos dados e auxiliando na identificação de temas mais frequentes para cada Ministério.

Por fim, a técnica de *topic modeling* foi empregada para identificar os principais tópicos presentes no *corpus*, por meio do algoritmo LDA (*Latent Dirichlet Allocation*). O LDA é um método estatístico que agrupa palavras em tópicos com base na sua co-ocorrência nos documentos (Blei, 2012) – neste caso, as manifestações. Esta abordagem é adequada para análises exploratórias, uma vez que sua estrutura não-supervisionada dispensa que os dados possuam classes *a priori*, como era o caso desse estudo, resultando em um conjunto de tópicos mais coerente e mais próximo do julgamento humano (Marcolin *et al.*, 2023; Nikolenko *et al.*, 2017). Para esta análise, foi utilizado pacote *pyLDAvis* em Python, que não apenas executa o modelo, mas também oferece uma visualização interativa dos tópicos, facilitando a interpretação dos resultados. A ferramenta permite explorar a distribuição das palavras em cada tópico e sua relevância, além de visualizar a proximidade entre os temas identificados.

A aplicação do *pyLDAvis*, assim como todas as outras técnicas, foi realizada de forma independente por Ministério e a visualização interativa ajudou a compreender a composição de cada tópico, destacando as palavras mais relevantes e sua distribuição ao longo do *corpus*. Essa análise permitiu uma categorização mais ajustada, fornecendo subsídios para uma elaboração de classes e categorias baseadas nos dados.

## 4. RESULTADOS OBTIDOS E CONTRIBUIÇÕES

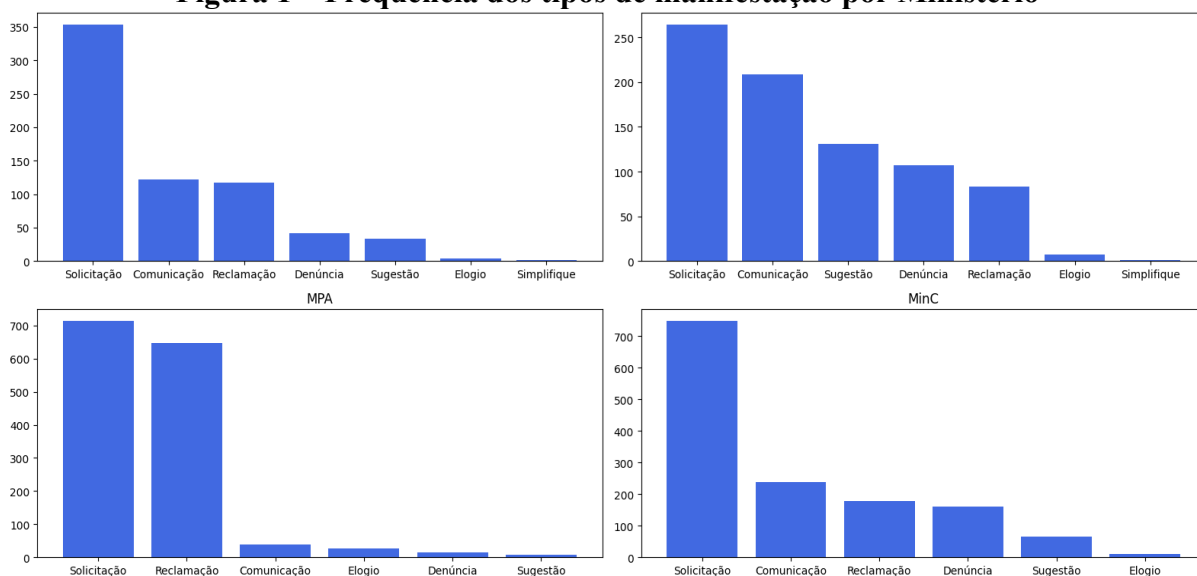
### 4.1 Exploração inicial

O intuito da primeira análise exploratória foi buscar uma visão geral dos dados estruturados das manifestações, incluindo tipo de manifestação, distribuição geográfica e categorização temática (pré-definida).

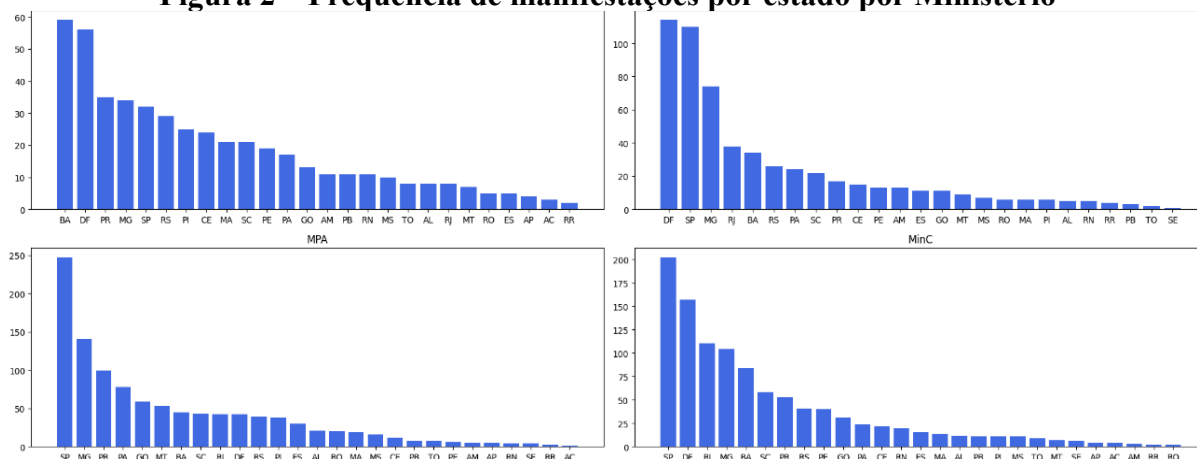
As manifestações podem ser classificadas em seis tipos diferentes: comunicação, denúncia, elogio, reclamação, solicitação e sugestão. Na Figura 1, é apresentado o gráfico de barras com a frequência de cada categoria por Ministério.

A distribuição geográfica das manifestações (Figura 2) revela padrões específicos por Ministério: no MDA, predominam manifestações da Bahia e Distrito Federal; no MMA, concentram-se no Distrito Federal, São Paulo e Minas Gerais; no MPA, destacam-se as de São Paulo; e no MinC, prevalecem São Paulo e Distrito Federal.

**Figura 1 – Frequência dos tipos de manifestação por Ministério**



**Figura 2 – Frequência de manifestações por estado por Ministério**



Também foi feita uma análise dos principais assuntos abordados nas manifestações em cada Ministério, de acordo com assuntos pré-cadastrados pelas ouvidorias. No entanto, o resultado revelou limitações significativas. No MDA e MMA, a alta diversidade temática com diferentes hierarquias dificultou a interpretação. Já no MPA, houve foco em um assunto específico (“pesca amadora”), enquanto no MinC, os assuntos foram tão diversos que se acumularam na categoria “outros”.

Essa análise preliminar das manifestações demonstrou que a utilização direta dos dados, sem tratamento adequado, pode não ser suficiente para compreender as principais demandas. A diversidade hierárquica dos assuntos pré-cadastrados, por exemplo, pode afetar a interpretação dos dados, enquanto a presença de dados repetidos ou não representativos de pedidos efetivos poderia desviar o foco da análise. Sendo assim, na tentativa de se obter resultados mais robustos e significativos, optou-se por realizar uma análise alternativa realizando um pré-processamento nos dados não-estruturados, o que permitiu uma exploração mais aprofundada e uma interpretação mais precisa dos resultados.

#### 4.2 Análise das manifestações aplicando critérios de exclusão

Para realizar esse tratamento nos dados, foram estabelecidos dois critérios de exclusão: remoção de comentários repetidos e eliminação de manifestações que não constituíam pedidos

efetivos. Esse processamento resultou em 4.168 manifestações válidas, distribuídas conforme apresentado na Tabela 2.

**Tabela 2 – Manifestações por Ministérios após a execução dos critérios de exclusão**

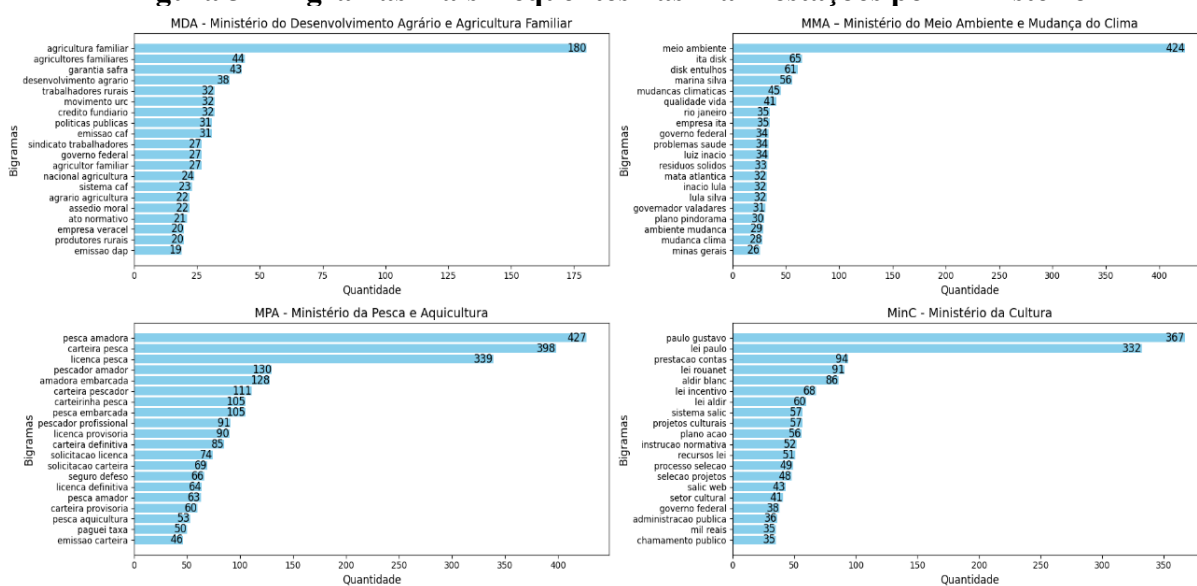
	Base de dados original	Base de dados após a limpeza
Ministério do Desenvolvimento Agrário e Agricultura Familiar	673	641
Ministério da Cultura	1.400	1.351
Ministério do Meio Ambiente e Mudança do Clima	802	754
Ministério da Pesca e Aquicultura	1.451	1.422
<b>TOTAL</b>	<b>4.326</b>	<b>4.168</b>

A redução do volume original (de 4.326 para 4.168 manifestações) foi relativamente pequena, mas permitiu uma melhor precisão analítica. Uma vez pré-processados, foram então aplicadas as técnicas de Processamento de Linguagem Natural aos dados de texto: análise de *n-grams*; etiquetagem *Part-of-Speech*; e *topic modeling*.

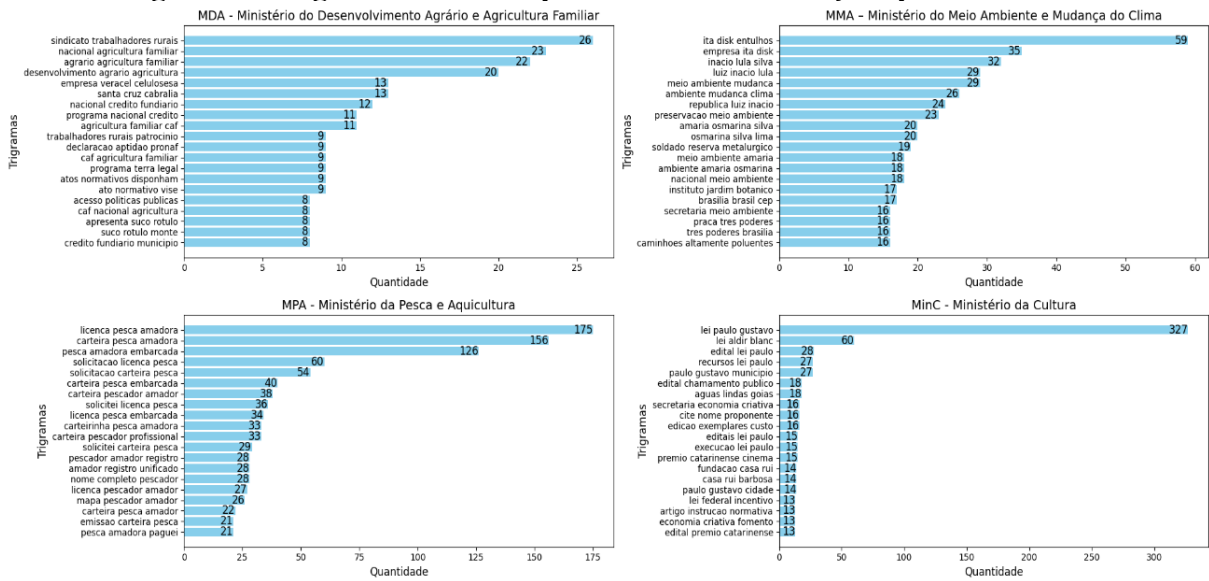
As análises semânticas (bigramas, trigramas e etiquetagem) evidenciam os termos mais associados a cada Ministério, permitindo um entendimento mais aprofundado do vocabulário utilizado e dos padrões de associação entre palavras. Essa abordagem possibilita capturar as principais demandas, preocupações e expectativas presentes nas manifestações recebidas pelas ouvidorias. Ao identificar essas nuances, é possível compreender melhor os temas recorrentes, as prioridades dos cidadãos e os desafios específicos de cada área, contribuindo para uma gestão mais assertiva e alinhada às necessidades da sociedade.

Nas Figuras 3 e 4, são apresentados, respectivamente, os bigramas e trigramas mais frequentes nas manifestações por Ministério.

**Figura 3 – Bigramas mais frequentes nas manifestações por Ministério**



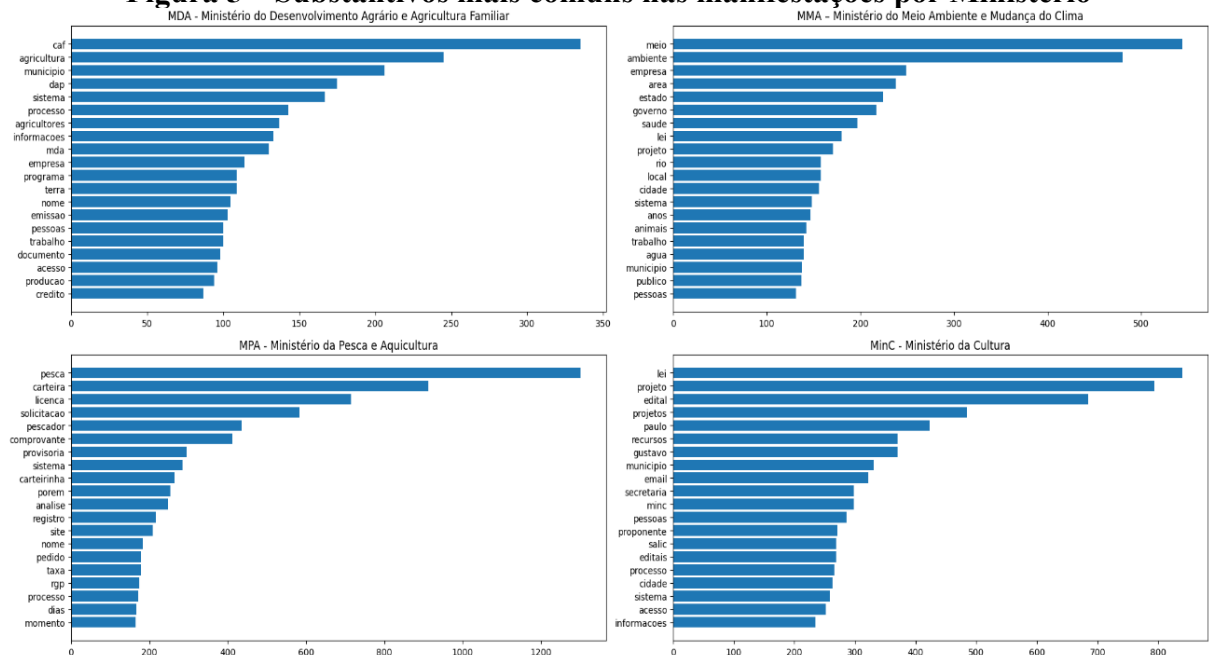
**Figura 4 – Trigramas mais frequentes nas manifestações por Ministério**



No Ministério do Desenvolvimento Agrário e Agricultura Familiar, os *n-grams* se concentram em temas de agricultura familiar, acesso a crédito fundiário, preocupações com sindicatos e políticas públicas para trabalhadores rurais. No Ministério do Meio Ambiente e Mudança do Clima, as manifestações abordam descarte irregular de entulhos, mudanças climáticas, preservação do meio ambiente e figuras públicas associadas ao Ministério. No Ministério da Pesca e Aquicultura, há um grande volume de manifestações sobre licenciamento e regulamentação da atividade pesqueira. No Ministério da Cultura, as demandas se concentram na Lei Paulo Gustavo, Lei Aldir Blanc, na Fundação Casa Rui Barbosa e em editais de fomento, destacando o interesse no acesso a recursos e incentivos culturais.

Na Figura 5, são apresentados os substantivos mais frequentes nas manifestações por Ministério.

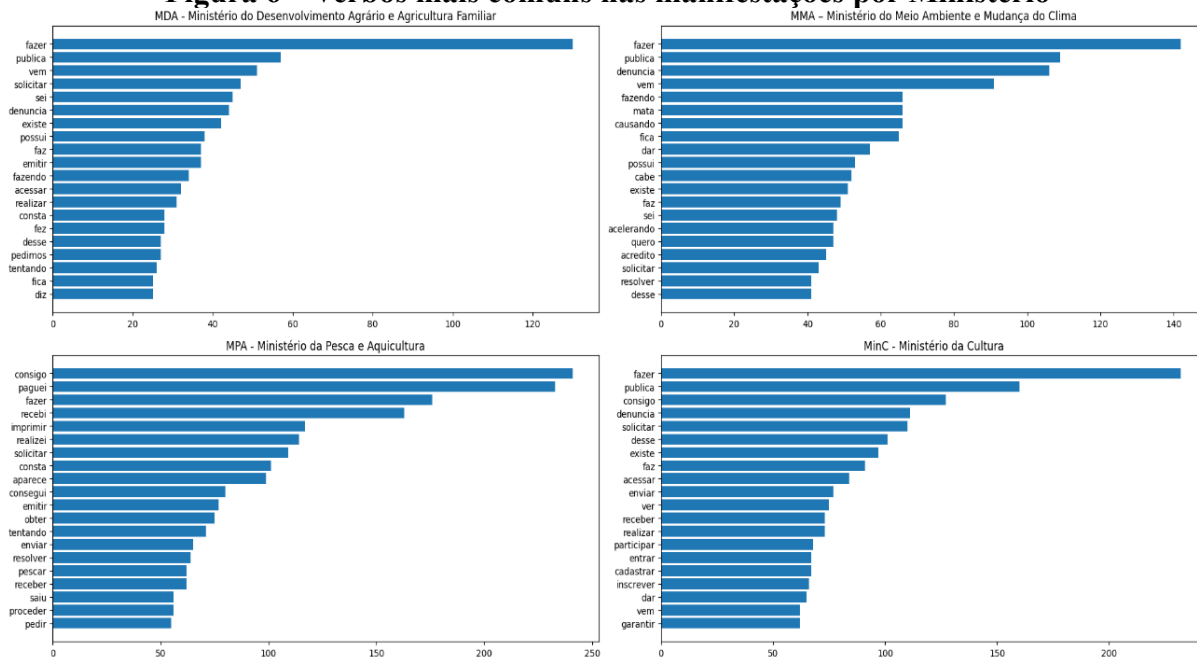
**Figura 5 – Substantivos mais comuns nas manifestações por Ministério**



No Ministério do Desenvolvimento Agrário e Agricultura Familiar, termos como “caf”, “dap” e “sistema” sugerem interesse em regulamentação junto ao Ministério, enquanto os termos "agricultura", "programa" e "crédito" apontam preocupações com acesso a políticas agrícolas e financiamento. No Ministério do Meio Ambiente e Mudança do Clima, palavras como “empresa”, "lei" e "projeto" indicam demandas sobre impactos ambientais e regulamentações. No Ministério da Pesca e Aquicultura, os substantivos mais frequentes demonstram que as principais manifestações estão relacionadas à regulamentação da atividade pesqueira. Já no Ministério da Cultura, a presença dos substantivos "lei", "projeto", "edital" e “recursos” sugere forte interesse em políticas de fomento cultural e acesso a recursos via editais.

Na Figura 6, são apresentados os verbos mais frequentes nas manifestações por Ministério.

**Figura 6 – Verbos mais comuns nas manifestações por Ministério**



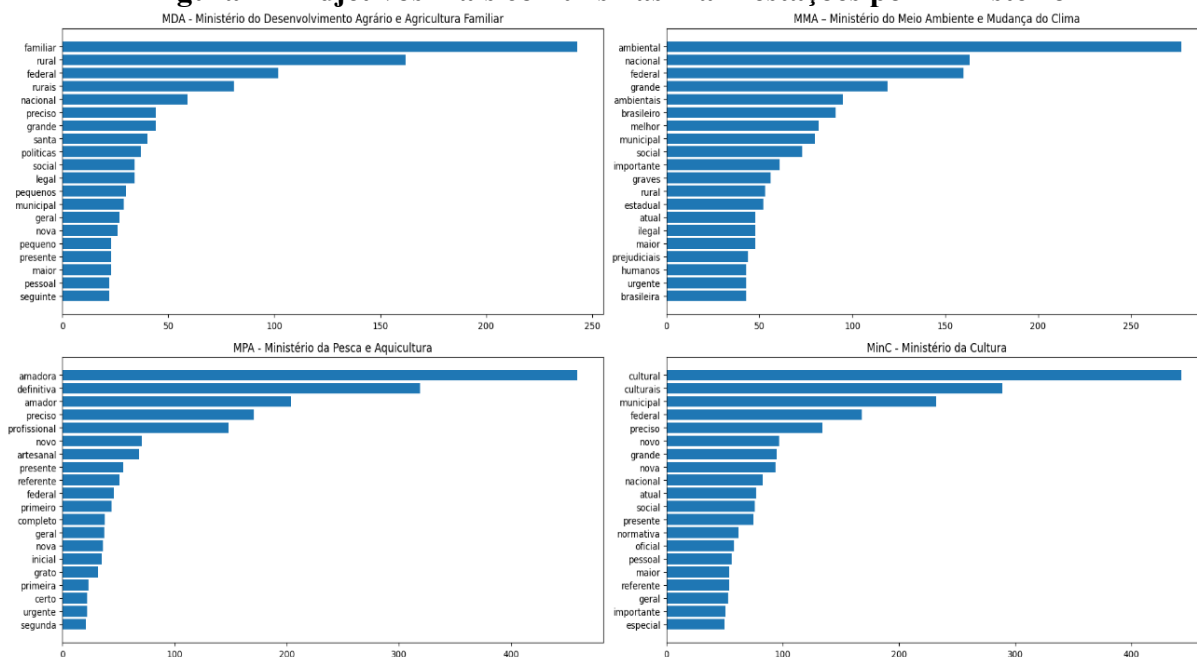
No Ministério do Desenvolvimento Agrário e Agricultura Familiar, verbos como "fazer", "solicitar", “pedir” e "denunciar" indicam tanto pedidos quanto queixas. No Ministério do Meio Ambiente e Mudança do Clima, a predominância de "denunciar", "causando", “mata”, “acelerando” e "resolver" sugere preocupações com impactos ambientais e demandas por ação governamental. No Ministério da Pesca e Aquicultura, os verbos se relacionam com o acesso a licenças e serviços e alguns em particular, como “tentando”, “resolver” e “proceder” podem indicar dificuldades nesses processos. Já no Ministério da Cultura, a presença de "acessar", "participar", "entrar" e “inscrever” aponta para interesse em participação cultural e acesso a recursos públicos.

Na Figura 7, são apresentados os adjetivos mais frequentes nas manifestações por Ministério.

No Ministério do Desenvolvimento Agrário e Agricultura Familiar, termos como "familiar", "rural" e "pequeno" destacam o foco nas demandas da agricultura familiar e dos pequenos produtores. No Ministério do Meio Ambiente e Mudança do Clima, a presença de adjetivos como “melhor”, “importante”, “grave”, “ilegal”, “prejudicial” e "urgente" sugere preocupações com políticas ambientais e a necessidade de ações imediatas. No Ministério da Pesca e Aquicultura, adjetivos como "amadora", "profissional" e "definitiva" indicam distinções entre diferentes categorias de pescadores e questões regulatórias. Já no Ministério da Cultura,

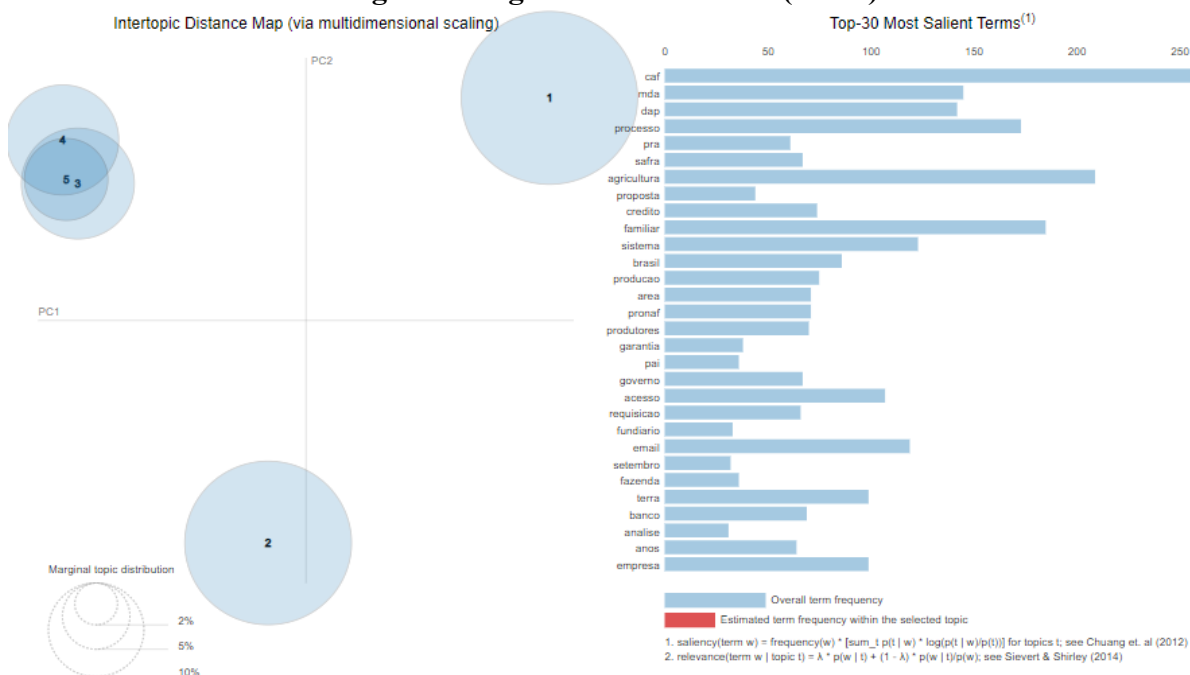
"cultural", "municipal" e "federal" apontam para manifestações relacionadas ao fomento cultural em diferentes esferas governamentais.

**Figura 7 – Adjetivos mais comuns nas manifestações por Ministério**



Na Figura 8, é apresentada a visualização da modelagem de tópicos das manifestações do Ministério do Desenvolvimento Agrário e Agricultura Familiar (MDA).

**Figura 8 – Modelagem de tópicos das manifestações do Ministério do Desenvolvimento Agrário e Agricultura Familiar (MDA)**



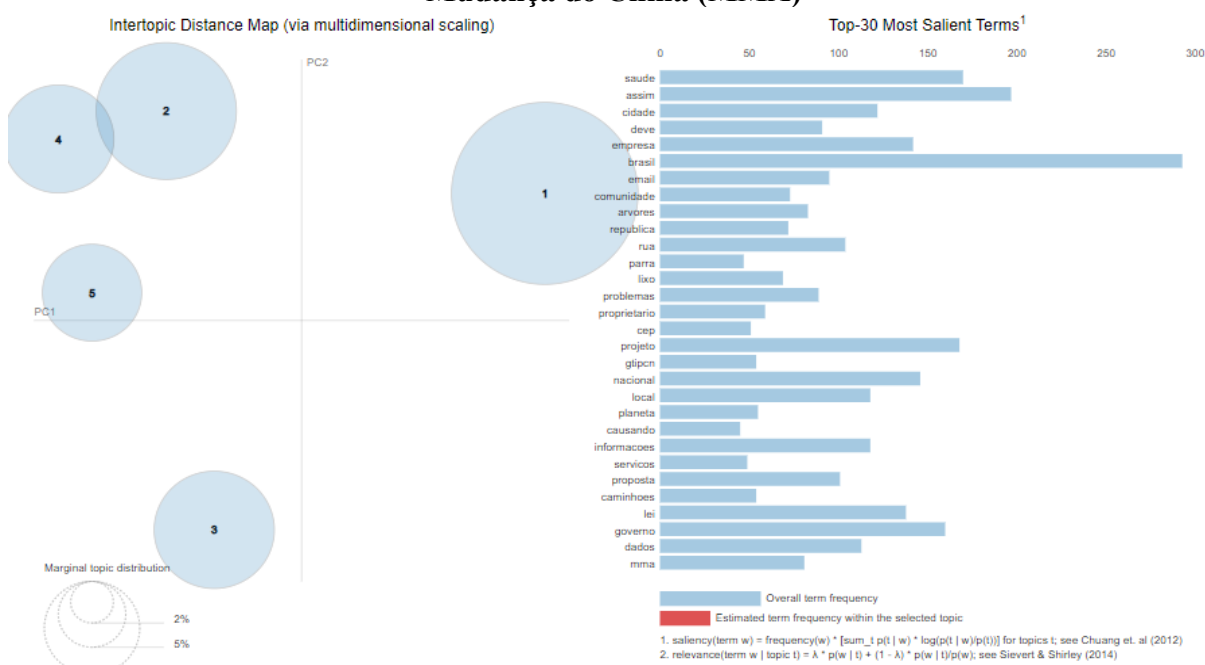
Com a modelagem de tópicos, o intuito foi visualizar como os termos presentes nas manifestações se agrupam em torno dos tópicos, representados pelos círculos, e quão relevantes

e interconectados esses tópicos são. O tamanho de cada círculo indica a relevância do tópico, enquanto a proximidade ou distância entre eles reflete a relação entre os temas. Quando os círculos se sobrepõem, há elementos linguísticos compartilhados entre os tópicos; já a distância entre eles demonstra a discricionariedade e a independência de cada assunto.

No caso do MDA, o tópico 1 reflete demandas burocráticas relacionadas aos processos administrativos do Ministério, envolvendo solicitações de documentos, prazos e problemas de acesso a sistemas (CAF, DAP, EMATER) e acesso a informações e serviços. O tópico 2 indica preocupações com políticas públicas, como programas de crédito (PRONAF) e apoio à agricultura familiar e pequenos produtores. Já os tópicos 3, 4 e 5 se sobrepõem, o que indica que estão fortemente relacionados. De maneira geral, os três tópicos abordam conflitos agrários, questões de regularização fundiária e programas de subsídio ou benefícios rurais.

Na Figura 9, é apresentada a visualização da modelagem de tópicos das manifestações do Ministério do Meio Ambiente e Mudança do Clima (MMA).

**Figura 9 – Modelagem de tópicos das manifestações do Ministério do Meio Ambiente e Mudança do Clima (MMA)**



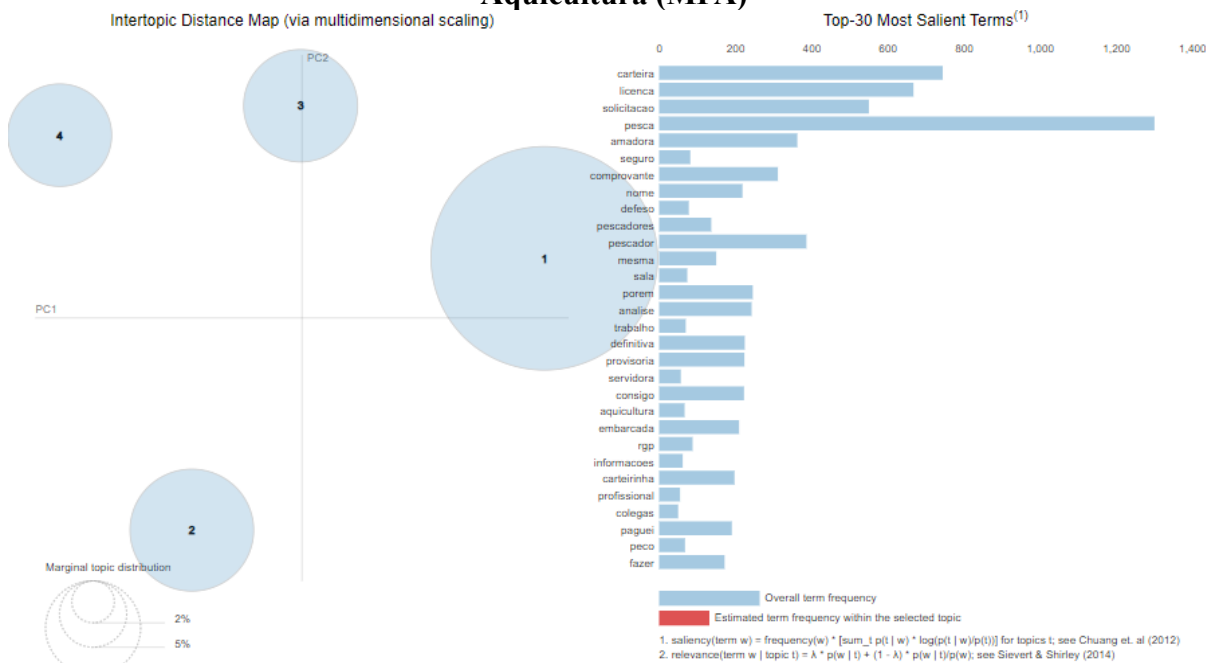
O tópico 1 apresenta termos que apontam para os temas de legislação ambiental, projetos ambientais e preservação de recursos naturais, refletindo uma preocupação com políticas públicas, planejamento e atuação do governo federal. O tópico 2 aborda conflitos em esferas locais, com destaque para atividades poluidoras e má gestão de resíduos. Já o tópico 3 apresenta termos associados às demandas burocráticas relacionadas aos processos administrativos do Ministério. O tópico 4, levemente relacionado ao tópico 2, está relacionado a denúncias e fiscalização, com foco em infrações ambientais. Por fim, o tópico 5 destaca os impactos ambientais, como danos à biodiversidade, e participação social.

Na Figura 10, é apresentada a visualização da modelagem de tópicos das manifestações do Ministério da Pesca e Aquicultura (MPA).

O tópico 1 reflete demandas burocráticas relacionadas aos processos administrativos do Ministério, como emissão de documentos, taxas e outros serviços. Os termos do tópico 2 apontam para questões relacionadas aos pescadores profissionais, como registro, seguro e direitos, sugerindo preocupações com políticas públicas e benefícios para a categoria. O tópico

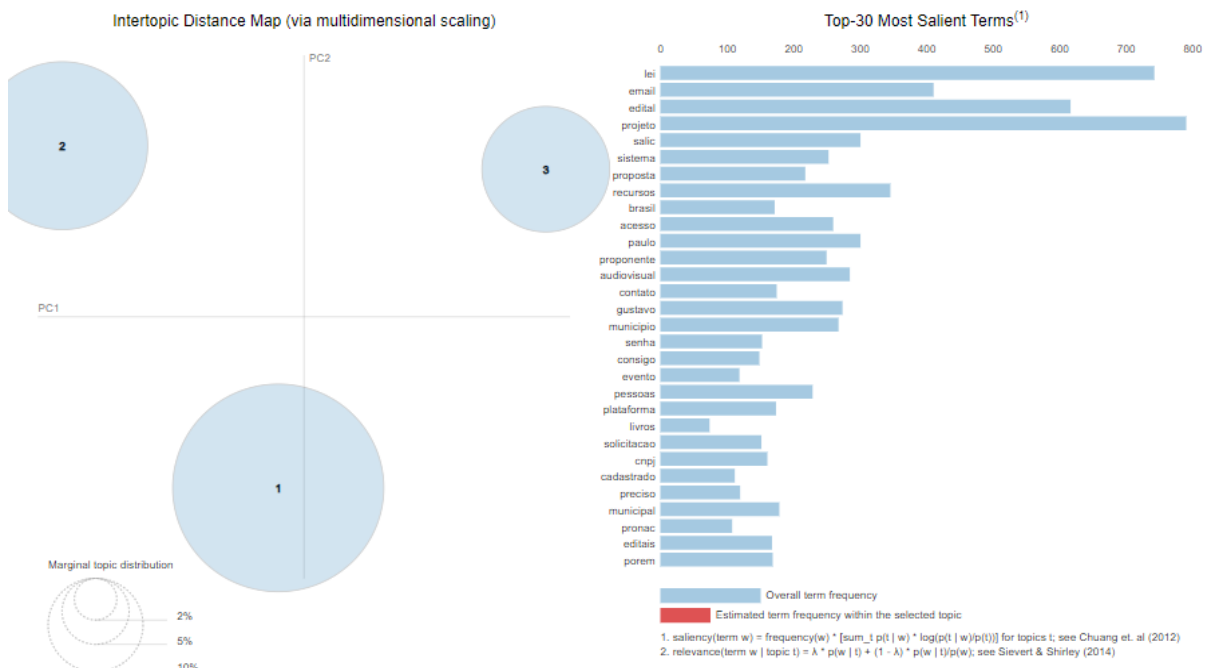
3 apresenta temas institucionais e legais envolvendo a atividade pesqueira. Por fim, o tópicos 4 parece apontar para conflitos internos, indicando manifestações sobre condições laborais e relações de trabalho dentro do próprio Ministério.

**Figura 10 – Modelagem de tópicos das manifestações do Ministério da Pesca e Aquicultura (MPA)**



A Figura 11 apresenta a visualização da modelagem de tópicos das manifestações do Ministério da Cultura (MinC).

**Figura 11 – Modelagem de tópicos das manifestações do Ministério da Cultura (MinC)**



O tópico 1 apresenta termos associados à gestão de recursos e projetos públicos, com foco em demandas de financiamento e políticas culturais locais. Já o tópico 2 reflete demandas burocráticas relacionadas aos processos administrativos do Ministério, envolvendo cadastros, dúvidas sobre editais/programas culturais e outras dificuldades. Por fim, o tópico 3 aborda a cultura e identidade nacional, destacando termos que envolvem diversidade cultural, eventos artísticos e outras produções.

\*\*\*

A análise comparativa dos resultados de cada técnica revela padrões específicos de interação entre os cidadãos e cada Ministério.

No Ministério do Desenvolvimento Agrário e Agricultura Familiar (MDA), as manifestações concentram-se em três eixos principais: questões burocráticas envolvendo sistemas administrativos, políticas públicas e programas de crédito para pequenos produtores, e conflitos agrários e regularização fundiária. O vocabulário predominante reflete um perfil de usuários que buscam tanto acesso a serviços quanto resolução de conflitos.

No Ministério do Meio Ambiente e Mudança do Clima (MMA), prevalecem temas relacionados ao descarte irregular de resíduos, licenciamento ambiental, preservação ambiental e políticas públicas para sustentabilidade. O caráter das manifestações é marcadamente denunciatório, o que mostra que este Ministério atende tanto demandas proativas (licenciamento, projetos) quanto reativas (denúncias, fiscalização).

No Ministério da Pesca e Aquicultura (MPA), as manifestações focam intensamente na regulamentação da atividade pesqueira, com destaque para a pesca amadora, licenciamento e registro de pescadores profissionais. O vocabulário técnico específico e a distinção clara entre categorias ("amadora", "profissional") indicam um público especializado e verbos como "tentando", "resolver" e "proceder" sugerem dificuldades nos processos burocráticos.

No Ministério da Cultura (MinC), as manifestações concentram-se no acesso a recursos culturais, especificamente na Lei Paulo Gustavo, Lei Aldir Blanc e outros editais de fomento. O perfil lexical evidencia um público interessado em participação ativa e financiamento cultural, o que indica um Ministério que opera primordialmente como facilitador de acesso a recursos públicos para produção cultural.

De maneira geral, todos os ministérios apresentam um tópico relacionado a "demandas burocráticas", indicando desafios comuns nos processos administrativos. Contudo, diferem significativamente no perfil dos usuários: o MDA atende principalmente produtores rurais individuais; o MMA recebe demandas tanto de cidadãos quanto de organizações; o MPA serve uma categoria profissional específica; e o MinC interage com um público culturalmente engajado. Essas especificidades sugerem a necessidade de estratégias de atendimento diferenciadas para cada contexto.

Para além das sugestões já mencionadas, ainda é possível fazer algumas recomendações gerais. Um primeiro ponto é a necessidade de as ouvidorias, sob coordenação e orientação da OGU, desenvolverem um modelo mais padronizado para a classificação dos assuntos das manifestações. Atualmente, os assuntos nem sempre são aderentes aos dados, com pouca comparabilidade entre as diferentes ouvidorias e ainda com vários pedidos que não são classificados. Com base nas análises, sugere-se elaboração de guia de classificação de assuntos que subsidie o trabalho das ouvidorias. Essa lista de assuntos fundamentada pode facilitar a classificação e possibilitar visões mais longitudinais dos dados. Uma segunda recomendação é de que as ouvidorias adotem os métodos aqui apresentados, com aplicação e análise periódica; adicionalmente, recomenda-se coleta de informações demográficas (como gênero e idade) para aprimorar as análises.

Ainda, as informações extraídas permitem compreender quais são padrões nas manifestações, que poderiam gerar: disponibilização de informações no site das organizações, em linguagem simples e de fácil acesso; respostas padronizadas, mas que atendam a grupos de demandas parecidos, cuidando para não haver uma padronização indiscriminada, visando atingir um equilíbrio entre padronização e customização; e modelos de respostas considerando o perfil da demanda, considerando que há ministérios com perfis de demanda vinda de organizações ou intermediários, e outros que têm demandas vindas de cidadãos-usuários.

A intervenção proposta apresenta avanços metodológicos para a análise de manifestações em ouvidorias públicas ao integrar sistematicamente múltiplas técnicas de Processamento de Linguagem Natural em uma abordagem unificada. Diferentemente de abordagens isoladas, combinar tais técnicas permite capturar tanto padrões lexicais quanto estruturas semânticas, identificando especificidades temáticas de cada Ministério que proporcionam visão mais completa das demandas cidadãs.

Além disso, a escalabilidade da metodologia é um diferencial importante, já que é possível processar de maneira automatizada grandes volumes de dados com recursos computacionais moderados. O uso de Python e de bibliotecas *open-source* garante a replicabilidade e a sustentabilidade da solução para diferentes contextos de ouvidorias públicas brasileiras.

Considerando que as técnicas trabalham com padrões, identificá-los pode subsidiar canais de comunicação mais efetivos, considerando o vocabulário e as preocupações específicas de cada área. De maneira geral, a metodologia permite capturar nuances das manifestações que talvez não fossem tão evidentes ou sequer apareceriam em análises tradicionais. Ainda, se incorporadas em um processo contínuo, além de permitir uma melhoria gradativa dos modelos pela base de conhecimento, essas análises possibilitam o acompanhamento da evolução das demandas ao longo do tempo, o que pode relevar tendências emergentes e permitir a antecipação de problemas e ajustes proativos nas políticas públicas, alinhados às necessidades sociais. Essa capacidade preditiva representa um avanço para a gestão pública responsiva.

## 5. CONSIDERAÇÕES FINAIS

Este artigo teve como objetivos apresentar uma metodologia para coleta, sistematização e análise dos dados de ouvidoria e, a partir disso, subsidiar a Ouvidoria Geral da União com recomendações para a melhoria da gestão das informações das ouvidorias.

Para a literatura, este estudo contribui ao demonstrar a viabilidade de aplicação de técnicas de Inteligência Artificial em contextos governamentais reais. A proposta metodológica oferece um modelo replicável para outras ouvidorias, de aumento gradativo de qualidade pelo uso continuado e alimentado de novas ocorrências, respeitando as especificidades institucionais e temáticas de cada órgão. É uma solução sustentável que representa uma alternativa viável para a Administração Pública, especialmente considerando limitações orçamentárias e a necessidade de transparência tecnológica.

Para a prática, a metodologia proposta proporciona um avanço na compreensão e gestão das manifestações dos cidadãos. Ao implementar técnicas de Processamento de Linguagem Natural, como neste estudo demonstradas, é possível criar um sistema mais eficiente e responsivo. A abordagem permite identificar padrões nas manifestações recebidas, possibilitando que gestores públicos antecipem necessidades, redesenhem políticas públicas de forma mais assertiva e estabeleçam canais de comunicação mais diretos e transparentes.

Essa metodologia evidencia que a transformação digital no setor público não precisa se limitar à digitalização e otimização de processos administrativos. Pode, também, proporcionar uma inteligência analítica que qualifica a interação entre Estado e sociedade, promovendo uma aproximação através de uma escuta mais qualificada, tecnologicamente sofisticada e responsiva às demandas sociais.

Por fim, a transparência algorítmica é fundamental neste contexto e em todo esse processo – os cidadãos precisam estar cientes e precisam ser capazes de compreender como suas manifestações são processadas e categorizadas, garantindo a legitimidade da intervenção aqui proposta. Esse compromisso com a transparência fortalece a democracia participativa e consolida o uso responsável da tecnologia na gestão pública.

## 6. REFERÊNCIAS

ALSHAHRANI, A.; DENNEHY, D.; MANTYMAKI, M. An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. **Government Information Quarterly**, v. 39, n. 4, p. 1-14. 2022. Disponível em: <https://doi.org/10.1016/j.giq.2021.101617>

BAHN, L.; STROBEL, G. Generative artificial intelligence. **Electronic Markets**, v. 33, n. 63, p. 33-63, dez. 2023. Disponível em: <https://doi.org/10.1007/s12525-023-00680-1>

BLEI, D. M. Probabilistic topic models. **Communications of the ACM**, v. 55, n. 4, p. 77-84. 2012. Disponível em: <https://doi.org/10.1145/2133806.2133826>

BREIT, E.; EGELAND, C.; LOBERG, I. B.; ROHNEBAEK, M. T. Digital coping: How frontline workers cope with digital service encounters. **Social Policy & Administration**, v. 55, n. 5, p. 833-847. 2020. Disponível em: <https://doi.org/10.1111/spol.12664>

CHOLYSHKINA, O.; KARNAUKH, A.; VOLIANIUK, O.; OSTAPENKO, M.; HOLISHEVSKA, A. Public participation and innovative technologies: the role of artificial intelligence in public administration and sustainable development. **Salud, Ciencia y Tecnología - Serie de Conferencias**, v. 3, n. 974, jan. 2024. Disponível em: <https://doi.org/10.56294/sctconf2024974>

CONTROLADORIA GERAL DA UNIÃO. **Manual de Ouvidoria Pública**. Brasília: Ascom/CGU, 2019. Disponível em: [https://www.gov.br/ouvidorias/pt-br/central-de-conteudos/biblioteca/arquivos/ManualdeOuvidoriaPublica2019\\_web.pdf](https://www.gov.br/ouvidorias/pt-br/central-de-conteudos/biblioteca/arquivos/ManualdeOuvidoriaPublica2019_web.pdf)

ECKHARD, S.; FRIEDRICH, L. Linguistic features of public service encounters: How spoken administrative language affects citizen satisfaction. **Journal of Public Administration Research and Theory**, v. 34, n. 1, p. 122–135. 2024. Disponível em: <https://doi.org/10.1093/jopart/muac052>

GESK, T. S.; LEYER, M. Artificial intelligence in public services: when and why citizens accept its usage. **Government Information Quarterly**, v. 39, n. 3, p. 1-18, jul. 2022. Disponível em: <https://doi.org/10.1016/j.giq.2022.101704>

GILAD, S.; FREEDMAN, M.; BLOOM, P. B. N. Citizens' Communication Styles in Written Public Encounters. **Perspectives on Public Management and Governance**, v. 7, n. 3, p. 75-88, 2024. Disponível em: <https://doi.org/10.1093/ppmgov/gvae004>

LINDGREN, I.; MADSEN, C. O.; HOFMANN, S.; MELIN, U. Close encounters of the digital kind: A research agenda for the digitalization of public services. **Government Information Quarterly**, v. 36, n. 3, p. 427-436, jul. 2019. Disponível em: <https://doi.org/10.1016/j.giq.2019.03.002>

MADUPATI, B. Madupati, Bhanuprakash, The Role of AI in the Public Sector: A Technical Perspective. **Journal of Artificial Intelligence & Cloud Computing**, v. 3, n. 4, p. 1-6, ago. 2024. Disponível em: [http://dx.doi.org/10.47363/JAICC/2024\(3\)E147](http://dx.doi.org/10.47363/JAICC/2024(3)E147)

MARCOLIN, C. B.; DINIZ, E. H.; BECKER, J. L.; DE OLIVEIRA, H. P. G. Who knows it better? Reassessing human qualitative analysis with text mining. **Qualitative Research in Organizations and Management**, v. 18, n. 2, p. 181-198. 2023. Disponível em: <https://doi.org/10.1108/QROM-07-2021-2173>

MERGEL, I.; EDELMANN, N.; HAUGA, N. Defining digital transformation: Results from expert interviews. **Government Information Quarterly**, v. 36, n. 4, p. 1-16. 2023. Disponível em: <https://doi.org/10.1016/j.giq.2019.06.002>

NIKOLENKO, S. I.; KOLTCOV, S.; KOLTSOVA, O. Topic modelling for qualitative studies. **Journal of Information Science**, v. 43, n. 1, p. 88-102. 2017. Disponível em: <https://doi.org/10.1177/0165551515617393>

RAAPHORST, N.; VAN DE WALLE, S. A signaling perspective on bureaucratic encounters: How public officials interpret signals and cues. **Social Policy & Administration**, v. 52, n. 7, p. 1367-1378. 2017. Disponível em: <https://doi.org/10.1111/spol.12369>

PEETERS, R. Digital Administrative Burdens: An Agenda for Analyzing the Citizen Experience of Digital Bureaucratic Encounters. **Perspectives on Public Management and Governance**, v. 6, n. 1, p. 7-13, mar. 2023. Disponível em: <https://doi.org/10.1093/ppmgov/gvac024>

POSTIGLIONE, A. Linguistic Text Mining. In: TÁLLON-BALLESTEROS, A. J. (Ed.) **Fuzzy Systems and Data Mining X**. Amsterdam, Netherlands: IOS Press, 2024, p. 314-331. Disponível em: <https://ebooks.iospress.nl/doi/10.3233/FAIA241433>

SARKAR, D. **Text Analytics with Python: A Practitioner's Guide to Natural Language Processing**. 2. ed. Bangalore: Apress, 2019. 688 p. Disponível em: <https://doi.org/10.1007/978-1-4842-4354-1>