



II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO

(X) CONSÓRCIO MESTRAL (CM) - 10 a 13 páginas

() CONSÓRCIO DOUTORAL (CM) – 12 a 15 páginas

APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA GESTÃO DA INFORMAÇÃO DAS DEMANDAS POR SERVIÇOS PÚBLICOS DA PREFEITURA DE CURITIBA

MACHINE LEARNING APPLICATION FOR INFORMATION MANAGEMENT OF DEMANDS BY PUBLIC SERVICES OF CURITIBA CITY HALL

APLICACIÓN DE APRENDIZAJE DE MÁQUINA PARA LA GESTIÓN DE LA INFORMACIÓN DE LAS DEMANDAS DE LOS SERVICIOS PÚBLICOS DEL AYUNTAMIENTO DE CURITIBA

Lucimara Wons

Mestranda no Programa de Pós-Graduação em Gestão da Informação

Universidade Federal do Paraná, Curitiba, PR, Brasil

 0000-0003-0123-6923 E-mail: luci@wons.com.br

Ricardo Mendes Junior

Doutor em Engenharia de Produção

Universidade Federal do Paraná, Curitiba, PR, Brasil

 0000-0003-4947-0364 E-mail: ricardomendesjr@gmail.com

Resumo: As centrais 156 de atendimento ao cidadão complementam os instrumentos de participação popular previstos na Constituição Federal, recebendo diariamente demandas, alusivas aos serviços públicos, geralmente em texto livre. O objetivo da pesquisa é aplicar algoritmos de aprendizado de máquina para classificação automática dessas demandas na Prefeitura de Curitiba. A primeira fase da pesquisa, caracterizada como descritiva, quantitativa e aplicada, teve o propósito de identificar os termos mais citados nas demandas e compará-los à estrutura de assuntos da prefeitura. O método utilizado compreendeu as cinco etapas do processo de mineração de textos: coleta, pré-processamento, indexação, mineração e análise dos dados, além da técnica de associação de termos. Como resultados dessa fase, foram obtidos 92 termos mais citados no conjunto de demandas, que indicaram 23 assuntos, sendo 19 existentes na tabela da prefeitura. Os assuntos mais citados foram “coleta”, “trânsito” e “abordagem social de rua”, totalizando 53,09% das citações. As considerações parciais apontam que há alinhamento entre os termos mais citados e os termos associados à tabela de assuntos da prefeitura. Também indicam que existem reclamações associadas à falta de informação, além daquelas já mapeadas pela prefeitura para as unidades de saúde. Essa questão revelada pela mineração é potencialmente útil para otimizar a comunicação com os cidadãos. A pesquisa visa contribuir com as discussões acerca do uso da inteligência artificial em prol do atendimento ao cidadão, colaborando com a gestão da informação na Prefeitura de Curitiba.

Palavras-chave: Mineração de textos. Técnica de associação. Gestão da Informação. Serviços municipais. Gestão pública.

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

Abstract: The 156 Citizen Service Centers are complementary to the instruments of popular participation provided by the Federal Constitution, receiving daily demands, allusive to public services, usually in free text. The objective of this research is to apply machine learning algorithms to automatically sort these demands in the Curitiba City Hall. The first phase of the research, characterized as descriptive, quantitative and applied, had the purpose of identifying the most quoted terms in the demands and compare these terms to the town hall's subject structure. The method used comprehended the five steps of the text mining process: data collection, pre-processing, indexing, mining and data analysis, in addition to the term association technique. As part of the result of this phase were obtained 92 most quoted terms in the demand set, which indicated 23 subjects and of those, 19 existed in the town hall's table. The most quoted subjects were "collect", "traffic" and "social street approach", totalizing 53,09% of all citations. The partial considerations point to an alignment between the most quoted terms and the terms associated with the town hall's table of subjects. They also indicate that there are complaints associated to lack of information, other than those already mapped by the town hall for health units. This matter revealed by the mining is potentially useful in optimizing the communication with citizens. The research aims to contribute with discussions about the use of artificial intelligence to help with citizen services, collaborating with the information management in Curitiba City Hall.

Keywords: Text mining. Association rules. Information Management. Public services. Public management.

Resumen: Los 156 Centros de Atención al Ciudadano son complementarios a los instrumentos de participación popular previstos por la Constitución Federal, recibiendo demandas diarias, alusivas a los servicios públicos, generalmente en texto libre. El objetivo de esta investigación es aplicar algoritmos de aprendizaje automático para ordenar automáticamente estas demandas en el Ayuntamiento de Curitiba. La primera fase de la investigación, caracterizada como descriptiva, cuantitativa y aplicada, tuvo como objetivo identificar los términos más cotizados en las demandas y comparar estos términos con la estructura temática del ayuntamiento. El método utilizado comprendió los cinco pasos del proceso de minería de texto: recolección de datos, preprocesamiento, indexación, minería y análisis de datos, además de la técnica de asociación de términos. Como parte del resultado de esta fase se obtuvieron los 92 términos más cotizados en el conjunto de demanda, que revelaron 23 sujetos y de ellos, 19 existían en la mesa del ayuntamiento. Los temas más citados fueron "recoger", "tráfico" y "acercamiento social a la calle", totalizando el 53,09% del total de citas. Las consideraciones parciales señalan que existe una alineación entre los términos más citados y los asociados con la tabla de materias del ayuntamiento. También indican que existen denuncias asociadas a la falta de información, además de las ya mapeadas por el ayuntamiento para las unidades de salud. Este problema revelado por la minería es potencialmente útil para optimizar la comunicación con los ciudadanos. La investigación tiene como objetivo contribuir a las discusiones sobre el uso de la inteligencia artificial en beneficio del ciudadano, colaborando con la gestión de la información en el Ayuntamiento de Curitiba.

Palabras clave: Minería de textos. Reglas de asociación. Gestión de la Información. Servicios públicos. Gestión pública.

1 INTRODUÇÃO

A informação é caracterizada como algo significativo para os indivíduos, geralmente interpretada a partir de uma representação simbólica contida em dados (SETZER, 1999, p. 2). A percepção da informação recebida, associada às aprendizagens anteriores e ao raciocínio, resulta na aquisição do conhecimento (SORDI, 2008, p. 9-12) que, no contexto organizacional, pode ser traduzido em maior eficiência e competitividade. Considerando-se que o volume de dados produzidos no mundo atinge 2,5 bilhões de gigabytes diários (SCHNEIDER, 2016, p. 2), torna-se fundamental para as organizações gerir dados e

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO

02 a 04 de setembro de 2020

informações. A gestão da informação auxilia indivíduos e organizações a acessar, processar e usar informações, consistindo no gerenciamento processos e sistemas que tratam e disponibilizam informações (DETLOR, 2010, p. 103-105).

Dados podem ser estruturados, como registros padronizados em bancos de dados, e não estruturados (FELDMAN; SANGER, 2007, p. 1), compreendendo documentos, relatórios, mensagens e vídeos, os quais correspondem a 80% dos dados produzidos no mundo (SCHNEIDER, 2016, p. 2). Mesmo que constituam uma fonte significativa de oportunidades de negócios para as organizações, tem sido um desafio transformar dados não estruturados em informação útil (CASTRO; FERRARI, 2016, p. 34; LAROSE; LAROSE, 2014, p. xi). Esse desafio mostra-se de modo análogo nas organizações públicas, haja vista o porte dessas instituições e o volume de dados que armazenam diariamente.

Dentre os dados não estruturados e acumulados pelas organizações públicas encontram-se demandas e contribuições que chegam por meio da participação popular. Millon (2010, p. 70) afirma que a participação popular é de suma importância para o atingimento das finalidades públicas, uma vez que os governantes precisam conhecer as necessidades, expectativas e carências sociais para sustentar a tomada de decisão política.

No Brasil, por meio da Emenda Constitucional nº 19 de 1988 a participação popular foi elevada à categoria de garantia constitucional, conferindo aos cidadãos e às associações representativas o direito legítimo de contribuir para formulação e implementação das políticas públicas pelo Estado (MENEZES, 2005, p. 2). A participação popular instituída confere ao usuário do serviço público o poder reclamatório sobre a qualidade dos serviços, a sua manutenção e sobre o acesso às informações dos atos dos governos (MENEZES, 2005, p. 2-3). A regulamentação do dispositivo constitucional só foi efetivada em junho de 2017, com a Lei nº 13.460/2017 que estabeleceu normas para a participação, proteção e defesa dos direitos do usuário dos serviços públicos, sejam esses prestados de forma direta ou indireta pela administração pública. Ponto importante, no artigo 5º, inciso XIII, a lei estabeleceu como diretriz a “aplicação de soluções tecnológicas que visem a simplificar processos e procedimentos de atendimento ao usuário e a propiciar melhores condições para o compartilhamento das informações” (BRASIL, 2017, p. 2).

O interesse pelo uso da tecnologia como agente de aumento da participação popular na gestão pública vem aumentando ano a ano, caracterizado pelo uso crescente das mídias sociais como meio de crítica ou apoio de grandes massas a fatos ocorridos. Porém, a participação popular necessita de uma estrutura de gestão da informação que permita à população ter acesso a informações sobre a cidade, além de serem ouvidas (MENDES JUNIOR, 2018, p. 2-3).

Solicitações, reclamações, sugestões e denúncias alusivas aos serviços públicos, em vários municípios brasileiros, são registradas por centrais de atendimento ao cidadão, que atendem pelo telefone 156. Número, esse, regulamentado pela Agência Nacional de Telecomunicações como de utilidade pública e específico para atendimento referente aos serviços municipais (ANATEL, 2015). As Centrais 156, como são nominadas, têm a função de estreitar e facilitar o relacionamento entre as prefeituras e os cidadãos, somando-se aos instrumentos de participação popular existentes.

Por meio da Central 156 de Curitiba, no ano de 2019, foram abertos cerca de 288 mil protocolos, em texto livre, referentes às demandas dos cidadãos (CURITIBA, 2019). Para o

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

encaminhamento aos órgãos competentes, os protocolos foram classificados manualmente, procedimento sujeito à falha humana. A gestão da informação, por meio da introdução de sistemas capazes de converter textos em dados estruturados, e posterior classificação automática, pode auxiliar na redução dessas falhas.

O aprendizado de máquina tem contribuído para classificação automática de dados e solução de problemas em várias áreas, como médica, jurídica e bancária. Na governamental tem sido utilizada para detecção de locais propensos a crimes e fiscalização de determinadas atividades econômicas, como sonegação de impostos e revisões de seguro social por invalidez (COGLIANESE; LEHR, 2019, p. 7-9),

Lopes *et al.* (2014, p. 3) definem o aprendizado de máquina como uma linha de pesquisa da inteligência artificial que tem por objetivo “estudar e desenvolver métodos computacionais para a obtenção de sistemas capazes de adquirir conhecimento de forma automática”. A técnica de classificação compõe o grupo de predições, termo que se refere à construção e uso de modelos computacionais capazes de prever a categoria, ou classe, de um elemento não rotulado (CASTRO; FERRARI, 2016, p. 41). Os autores explicam que, a partir de um grupo de elementos com classes ou rótulos conhecidos *a priori*, é possível treinar e ajustar um modelo, que efetue a classificação de elementos com classes ausentes.

Mehr *et al.* (2017, p. 1-2) destacam que os domínios da inteligência artificial, incluindo o aprendizado de máquina, podem ajudar a reduzir encargos administrativos e a resolver problemas de alocação de recursos, uma vez que os serviços ao cidadão compreendem questões como pesquisar e verificar o trâmite de documentos e elaborar documentos solicitados. Nesses casos, sistemas informatizados podem classificar as demandas e dar retorno aos cidadãos automaticamente, ou auxiliar os atendentes na classificação dos protocolos gerados.

É nesse contexto que se insere o problema do presente estudo: como o aprendizado de máquina pode auxiliar na classificação dos protocolos da Central 156 colaborando com a gestão da informação da Prefeitura de Curitiba?

O objetivo geral da pesquisa é aplicar algoritmos de aprendizado de máquina para classificação dos protocolos abertos via Central 156 de Curitiba. Os objetivos específicos são: a) identificar os termos mais frequentes no conjunto de protocolos; b) comparar esses termos à estrutura de assuntos da prefeitura; c) aplicar algoritmos do aprendizado de máquina num conjunto de exemplos rotulados para obtenção de um classificador; d) utilizar o classificador em novos exemplos, ainda não rotulados, e) avaliar o desempenho do classificador.

Neste artigo são apresentados resultados de estudo preliminar já realizado. Após esta introdução, a seção 2 traz definições acerca dos conceitos que sustentam a primeira fase da pesquisa. Na seção 3, são apresentados os materiais e métodos utilizados no desenvolvimento dessa fase, que contempla os objetivos a) e b). Na seção seguinte, relata-se os resultados parciais obtidos e as análises. Por fim, são expostas as considerações e as referências da pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Dentre os recursos da gestão da informação estão as aplicações de tecnologia (MARCHIORI, 2002, p. 75-76). No contexto atual, muitas dessas aplicações são voltadas ao

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

tratamento e processamento de volumes de dados que, a partir da identificação de padrões, extraem informações relevantes e auxiliam na tomada de decisão. O aprendizado de máquina e a mineração de textos compõem o rol dessas aplicações.

O aprendizado de máquina visa ao desenvolvimento de sistemas capazes de adquirir conhecimento de forma automática, ou seja, programas computacionais que tomam decisões baseadas em experiências (MONARD; BARANAUSKAS, 2003, p. 39; LOPES *et al.*, 2014, p. 4). Para Davenport e Prusak (1998, p. 4) o conhecimento não é algo claro, simples, sendo difícil compreendê-lo totalmente em termos lógicos. Nesse sentido, no caso do aprendizado de máquina, o conceito proposto pelos autores é congruente, ou seja, que o conhecimento é obtido com base nas experiências, informações e percepções dos indivíduos, incorporadas em documentos, repositórios, processos e práticas organizacionais (DAVENPORT; PRUSAK, 1998, p. 4).

No aprendizado de máquina, como no aprendizado humano, é utilizada a indução, ou referência lógica, para obter conclusões, derivando conhecimento novo a partir do conjunto de exemplos, contudo, se os exemplos forem insuficientes ou mal escolhidos, é possível que as hipóteses formuladas não sejam válidas (MONARD; BARANAUSKAS, 2003, p. 40). O aprendizado de máquina concentra-se em generalizar a experiência de treinamento para gerar hipóteses que conseguem prever resultados (KHANNA; AWAD, 2015, p. 1), como uma classe para um exemplo não rotulado

O conceito de aprendizado de máquina definido por Witten *et al.* (2017, p. xxiii) relaciona-se ao conceito da mineração de dados. Para os autores, a mineração tem por objetivo extrair informações implícitas, e até então desconhecidas, em conjuntos de dados, cabendo ao aprendizado de máquina fornecer as técnicas básicas para que a mineração possa ser realizada.

Monard e Baranauskas (2003, p. 40) explicam que no aprendizado supervisionado, uma das subdivisões do aprendizado de máquina, cada exemplo do conjunto de treinamento é representado por um vetor de características ou atributos (vetor de entrada) e a respectiva classe. Cabe ao algoritmo, dado um conjunto de exemplos de entrada, construir um classificador que consiga determinar as classes para os exemplos ainda não rotulados. Khanna e Awad (2015, p. 2) esclarecem que, geralmente, o classificador identifica a classe por inferência estatística. O classificador é também denominado hipótese ou descrição de conceito (MONARD; BARANAUSKAS, 2003, p. 40).

Um dos algoritmos de classificação mais empregados é o *Naïve Bayes* que utiliza o Teorema de Bayes para prever, estatisticamente, a probabilidade de um exemplo pertencer a uma classe. Nesse algoritmo, para facilitar os cálculos, assume-se que o valor de um atributo em uma determinada classe é independente dos valores dos demais atributos, na definição da classe (CASTRO; FERRARI, p. 299-300).

A avaliação do desempenho do classificador pode ser realizada por meio da matriz de confusão, que permite comparar a classificação prevista com a classificação real, exibindo o resultado nas seguintes combinações possíveis: falso positivo, verdadeiro positivo, falso negativo e verdadeiro negativo (KHANNA; AWAD, 2015, p. 2).

Em se tratando da mineração de textos, Feldman e Sanger (2007, p. 1) a definem como um processo intensivo de conhecimento, no qual são utilizadas ferramentas para análise de coleções de documentos. Semelhante à mineração de dados, que procura descobrir padrões

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

em dados estruturados, a mineração de textos procura tratar e extrair informações relevantes de textos, segundo determinado propósito (WITTEN *et al.*, 2017, p. 515).

Como os documentos não podem ser processados diretamente pelos classificadores e algoritmos de aprendizado de máquina, são transformados, durante a etapa de pré-processamento, numa representação estruturada de dados. Nessa estrutura, nominada *bag-of-words*, os documentos são representados por vetores e seus atributos (palavras ou termos) formam a dimensão da estrutura, recebendo peso 1 quando o documento contém o atributo e 0, para o contrário (FELDMAN; SANGER, 2007, p. 68).

A análise descritiva da mineração procura por padrões e relações entre os atributos, de modo a caracterizar e resumir o conjunto de dados (TAN *et al.* 2006, p. 7). A associação é a técnica da análise descritiva utilizada para descobrir padrões que denotam associações entre atributos (FELDMAN; SANGER, 2007, p. 25). Essa técnica extrai relações entre atributos de um conjunto de dados, no qual “se A então B”, indicando que quando o elemento A ocorre, o B tende a ocorrer (VASQUES *et al.*, 2017, p. 12). O algoritmo de Feinerer *et al.* (2008, p. 28) calcula as associações entre um determinado termo e os demais num conjunto de documentos, a partir de um parâmetro estipulado para correlação mínima.

3 PERCURSO METODOLÓGICO

A pesquisa é caracterizada como descritiva, de abordagem quantitativa e natureza aplicada. O estudo preliminar realizado foi conduzido com dados da Central 156 da Prefeitura de Curitiba, segundo o processo de mineração de textos proposto por Aranha e Passos (2006, p. 4), e que contempla cinco etapas: a coleta dos dados, o pré-processamento, a indexação, a mineração e a análise dos resultados.

A Central 156 da Prefeitura de Curitiba disponibiliza cinco canais para registro das demandas pelos cidadãos: o número telefônico 156, o sistema *web*, acessado pelo endereço <http://www.central156.org.br/>, o *chat* humano, o *chat* robô e o aplicativo *mobile* Curitiba 156. As demandas abertas via telefone e *chat* humano são registradas, em texto livre, pelos atendentes da Central. No sistema *web* o cidadão registra a demanda, também em texto livre que, posteriormente, passa por triagem e classificação pelos atendentes. As demandas são classificadas segundo assuntos e subassuntos, ou subdivisões, predefinidos. Quando uma demanda se refere a mais de um assunto, é aberto um protocolo para cada assunto, sendo registrado, no campo descrição, o texto referente à demanda. Em 2019 havia 242 assuntos, divididos em 780 subdivisões (CURITIBA, 2019).

Na etapa de coleta, foram selecionados, de modo aleatório, dados referentes aos meses março, julho e novembro de 2019, disponibilizados no Portal de Dados Abertos da Prefeitura, totalizando 74.248 protocolos. Dos 21 campos existentes, escolheu-se apenas dois para o processo de mineração, o campo “solicitacao”, relativo ao código do protocolo e o “descricao”, relativo à descrição da demanda.

Na etapa de pré-processamento, realizou-se a remoção de outliers ou ruídos, isso é, dados que não seguem o comportamento ou as características dos demais (CASTRO; FERRARI, 2016, p. 44), resultando em 69.704 registros. Com auxílio do software de código aberto R, foram utilizados recursos do pacote *tm*, primeiro para conversão do conjunto de textos num *corpus*, com tratamento para o idioma português, e depois para limpeza, que

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

incluiu: alteração dos caracteres para minúsculo; remoção de caracteres especiais e pontuação; remoção de números; substituição dos caracteres acentuados do idioma português; remoção das *stop words*, isto é, palavras irrelevantes como artigos e preposições, além de palavras do domínio do 156 com pouca significância e grande quantidade de repetições e; *stemming*, para redução das palavras à sua raiz.

Na etapa de indexação, após seleção das palavras mais significativas, o *corpus* foi representado por meio da matriz de termos e documentos ou, no inglês, *term document matrix* (TDM), que é baseada no modelo *bag-of-words*. Para Feinerer *et al.* (2008, p. 10), a TDM pode ser compreendida como uma estrutura que indexa documentos (linhas da matriz), termos (dimensões) e valores, calculados pela quantidade de vezes que os termos aparecem nos documentos, melhor dizendo, sua frequência ou, no inglês, *term frequency* (TF).

Na avaliação da tabela de serviços do 156 constatou-se que a nominação dos assuntos é formada, substancialmente, por duas palavras adjacentes, formando assim bigramas. Em seguida foram calculadas as frequências para todos os termos da TDM, sendo selecionados os 100 termos que apresentaram as maiores frequências entre todos os documentos.

A etapa da mineração teve por objetivo a descoberta de termos que apresentavam maior correlação aos 100 termos mais frequentes na TDM. Para tanto, utilizou-se a função *findAssocs()*, do pacote *tm* do R. Essa função trabalha com três parâmetros: a TDM, os termos chave, para o quais se deseja encontrar os termos correlacionados, e o limite mínimo de correlação, que deve ser um valor numérico entre 0 e 1. O cálculo da correlação é realizado para cada documento que contenha o termo pesquisado em relação aos demais termos desse documento, e os documentos sem o termo da pesquisa são ignorados. Um limite de correlação de 0,5 retorna termos com coocorrência de, pelo menos, 50% no conjunto de textos.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Os 100 termos mais frequentes totalizaram 108.015 citações. Desses, oito termos foram eliminados por não apresentarem conteúdo informacional à análise. Os 92 termos resultantes, e apresentados na Figura 1, foram agrupados segundo 19 assuntos da prefeitura. Foram estabelecidos quatro novos assuntos para acumular termos condizentes a assuntos diversos: “via pública”, “reclamações e elogios”, “recadastro atendimento” e “urgência”.

Os assuntos “coleta”, “trânsito” e “abordagem social de rua” totalizam 53,09% das citações existentes e os 10 termos mais frequentes “fiscalização trânsito”, “coleta lixo”, “lixo vegetal”, “manutenção luminária”, “via pública”, “abordagem pessoa”, “coleta entulho”, “entulho diverso”, “coleta calça” e “posto saúde” correspondem a 56,15% do total de citações.

O resultado da aplicação da técnica de associação permitiu uma análise entre os termos mais citados e os termos a estes correlacionados. Os limites mínimos de correlação dependem do número de documentos existentes e do número de termos de cada documento, portanto, variam para cada termo mais citado. Foram considerados os termos associados com maiores valores de correlação. Nesta análise, além do agrupamento dos termos por assunto, considerou-se a área à qual cada assunto pertence.

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO

02 a 04 de setembro de 2020

35%, por queima de lixo 31%, ocasionada por chaminé 28%, referente a cheiro, com 28%, fogão a lenha 24%, queima a céu aberto 23%, muita fumaça 21% e queima de papel, com 20%.

Os resultados obtidos para a área de Trânsito incluem os assuntos trânsito, semáforo e fiscalização de cabos, resumidos no Quadro 2.

Quadro 2 - Termos correlacionados aos termos mais citados da área de Trânsito

Assunto	Termo citado	Termo correlacionado (correlação)
Trânsito	fiscalização trânsito, trânsito veículo, veículo estacionado	estacionado calçada (0,23), estacionado irregularmente (0,21), guia rebaixada (0,16), estacionado faixa (0,13), remoção guincho (0,12), vaga moto (0,11), impedindo saída (0,10), alerta minuto (0,10), carga descarga (0,10)
	veículo abandonado, abandonado via fiscalização veículo	público veículo (0,45), abandonado local (0,25), abandonado meses (0,17), carcaça veículo (0,13), determinadas marcas e cores de veículos (0,12), furtado desmontado (0,09)
	bloqueio pista, fiscalização bloqueio	pista veículo (0,22), pista caminhão (0,20), pista cone (0,13), pista festa (0,10), lombada terra (0,10), pista caçamba (0,10)
	risco acidente	dirigir inadequadamente (0,38), acidente motorista (0,30) causando risco (0,19), alta velocidade (0,15), forma brusca (0,13), além imprudência (0,11), realizava ultrapassagem (0,11)

Fonte: Elaborado pelos autores (2020)

Os termos mais frequentes do assunto trânsito foram divididos em quatro grupos, sendo que fiscalização engloba duas subdivisões da tabela da prefeitura, a fiscalização imediata e o estacionamento irregular. Quanto às demandas por fiscalização, as maiores queixas são referentes a veículo estacionado na calçada, com 23% de correlação, estacionado irregularmente 21%, na guia rebaixada 16%, na faixa 13%, em vaga de moto 11%, impedindo a saída 10%, vaga de carga descarga 10%, vaga rotativa 10% e demandas para remoção por guincho 12%.

Na área da Assistência Social os termos mais citados apresentam uma correlação recíproca de 98% que pode ser interpretada como “abordagem social para pessoas morando ou dormindo na via pública”. Os termos correlacionados desta área identificam: a) o perfil da população, são senhores e homens, com 36% de correlação, mulheres 31%, crianças 26%, pessoas que parecem perdidas e/ou desorientadas 24%, alcoolizadas 16%, rapazes 16%, esmolando 14%, mulheres com criança de colo 14%, senhoras 12%, casais 10%, sob efeito de drogas 10%, em grupo 8% e moças, com 8%; b) os locais de abordagem, sendo morando ou dormindo embaixo de marquises, com 8% de correlação, e em locais específicos, como a Rodoferroviária, com 7% e; c) que os próprios moradores de rua solicitam abordagem para si, com 38% de correlação.

Os termos frequentes obtidos para a área de Obras Públicas são iluminação, pavimentação e drenagem. No resumo do Quadro 3, assunto iluminação, observa-se que os três últimos termos da lista evidenciam a preocupação da população com a questão da segurança, o que vem a confirmar os resultados obtidos por Maia *et al.* (2019, p. 12) que correlacionam o número de demandas de iluminação pública aos números de vítimas de homicídios dolosos e de roubo e lesão corporal, seguidos de morte. Quanto ao assunto pavimentação, os termos mais citados indicam essencialmente demandas para vias com

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO

02 a 04 de setembro de 2020

buracos 22%, implantação de asfalto definitivo e pavimentação de toda extensão da via, com 14% de correlação.

Quadro 3 - Termos correlacionados aos termos mais citados da área de Obras Públicas

Assunto	Termo citado	Termo correlacionado (correlação)
Iluminação pública	manutenção luminária, lâmpada queimada, e troca luminária, luminária lâmpada	queimada poste (0,35), fica oscilando (0,17), fica acesa (0,16), identificação visível (0,15), lâmpada quebrada (0,14), praça luminária (0,12), vias específicas (0,10), comércios específicos (0,09), local ermo (0,09), outro delito (0,09), escuridão trecho (0,06)
	manutenção pavimento, revitalização pavimento	pavimento com buraco (0,22), tapa buraco (0,12), vias específicas (0,12), muito buraco (0,09) e buraco grande (0,08)
Pavimentação	revitalização pavimento	asfalto definitivo (0,14), pavimento toda (0,14), tapa buraco (0,13), asfalto novo (0,09), muito remendo (0,09) passa ônibus (0,08)
	implantação asfalto	asfalto definitivo (0,14)

Fonte: Elaborado pelos autores (2020)

O Quadro 4 apresenta assuntos e termos da área da Saúde. Os termos do assunto elogios e reclamações foram agrupados no quadro por apresentarem muitas menções à saúde. Diferente das demais áreas, na saúde são os termos correlacionados que representam as subdivisões dos assuntos na tabela da prefeitura.

Quadro 4 - Termos correlacionados aos termos mais citados da área da Saúde

Assunto	Termo citado	Termo correlacionado (correlação)
Posto de saúde	elogia atendimento, toda equipe	elogia equipe (0,66), parabeniza equipe (0,27), bem atendido (0,22), elogia atendimento (0,18), equipe ótima (0,18), parabeniza todos (0,18), ótimo atendimento (0,17), agradece equipe (0,13) agradece dedicação (0,13), atendimento enfermeira (0,12), unidade saúde (0,12), atendimento médico (0,11), atendimento posto (0,10), recepção posto (0,09)
	posto saúde, unidade saúde, reclama atendimento	reclama posto (0,38), unidades específicas (0,25), saúde básica (0,23), pede averiguação (0,23), agendamento consulta (0,16), comprar injeção (0,14), falta medicamento (0,14), agilidade atendimento (0,11), aplicação vacina (0,11), informação necessita (0,11), confusa devido (0,11)
Reclamações e elogios	reclama falta	falta médico (0,31), falta medicamento (0,27), falta vacina (0,18), falta glicemia (0,17), informação falta (0,16), vacina meningocócica (0,12), falta divulgação (0,09), insulina posto (0,08)
	reclama demora	demora consulta (0,21), demora agendar (0,16), atendimento unidade (0,15), demora internamento (0,14), atendimento consulta (0,13), agendamento exame (0,12), demora entrega (0,14), atendimento médico (0,13), demora marcação (0,11), consultas especializadas como oftalmologia, neurologia (0,07), demora procedimento (0,06)

Fonte: Elaborado pelos autores (2020)

Os resultados obtidos confirmam a abertura dada pela prefeitura para que os cidadãos possam registrar situações ocorridas junto às unidades de saúde, como forma de monitorar e aprimorar a qualidade do atendimento prestado à população. Desse modo, as subdivisões para

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO

02 a 04 de setembro de 2020

os assuntos ligados à área contemplam praticamente todos os problemas que possam ocorrer, do quadro de funcionários à falta de material. A falta de recursos refere-se, com maiores valores de correlação, à falta de médicos, com 31%, de medicamentos 27%, vacina 18% e fita de glicemia, com 17%. Nesse item, também aparecem os termos falta de informação, com 15% e falta de divulgação, com 9%, indicativo de que ações referentes à comunicação entre os profissionais das unidades de saúde e os cidadãos podem ser reforçadas.

Os assuntos da área de Urbanismo incluem grande parte das citações referentes à fiscalização. Para o assunto fiscalização de terrenos baldios ou edificados os termos mais frequentes formam dois grupos condizentes às suas subdivisões: a) fiscalização e limpeza de terrenos, tendo como termos correlacionados, terreno baldio, com 20%, mato alto 16%, utilizado por andarilho 11% e uso de droga 10%; b) guia rebaixada, correlacionado a rebaixada irregular 33%, veículo estacionado 28% e rebaixada indevidamente 19%. Constatase, a partir dos termos correlacionados, que há preocupação da população com o uso indevido desses terrenos.

Quanto ao assunto passeio, os termos mais citados apresentam uma correlação recíproca de 98% que pode ser interpretada como “fiscalização de obstrução de calçada impedindo a passagem”. Os termos associados são passagem de pedestre, com 20% de correlação, condomínio instalou 20%, lixeira fixa 17%, passagem caçamba 16%, passagem morador 14%, reconstrução de calçada 12% e legislação deveria 11%, em referência à Lei Municipal 11.095 de 2004 (CURITIBA, 2004) que estabelece, no artigo 142, que toda edificação deve possuir no interior do lote depósito, isto é, lixeira para guarda provisória de resíduos, em local que facilite o acesso pelos coletores.

5 CONSIDERAÇÕES PARCIAIS

A dissertação, em elaboração, procura demonstrar como o aprendizado de máquina pode auxiliar na classificação dos protocolos da Central 156 de Curitiba. Buscou-se, por meio dos dois primeiros objetivos, verificar quais eram os termos mais citados no conjunto de demandas e se havia alinhamento desses com os assuntos e as subdivisões da tabela da prefeitura. A análise dos 92 termos mais citados indicou 23 assuntos, sendo 19 existentes na tabela da prefeitura.

Quanto aos demais resultados, na área de obras públicas, as demandas por pavimentação estão correlacionadas, em especial, a vias com buracos, asfalto definitivo e pavimentação em toda a extensão da via. A obstrução de calçadas, seja por veículo estacionado, por instalação de lixeiras ou caçambas, e a segurança são questões que aparecem em mais de um assunto. Na saúde, a falta de recursos está correlacionada, em especial, a médicos, medicamentos e vacinas.

As considerações parciais apontam que tanto os termos mais citados, quanto os termos correlacionados, estão alinhados aos assuntos e às subdivisões da tabela, indicando que o objetivo geral da dissertação, a classificação das demandas com aprendizado de máquina, encontra caminho favorável, uma vez que utilizará como “classe” os assuntos e as subdivisões. Também apontam que, além das reclamações já mapeadas pela prefeitura para as unidades de saúde, existem reclamações associadas à falta de informação. Essa informação

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO 02 a 04 de setembro de 2020

é potencialmente útil para fortalecimento das ações de comunicação dos profissionais da área com os cidadãos.

Quanto aos demais objetivos específicos da pesquisa: aplicar algoritmos de aprendizado num conjunto de exemplos rotulados para obtenção de um classificador, utilizar o classificador em exemplos não rotulados e avaliar o desempenho do classificador, pretende-se realizar experimentos e análises com estatística probabilística, por meio do software R.

A pesquisa procura contribuir com as discussões acerca do uso da inteligência artificial em prol do atendimento ao cidadão, no âmbito da Prefeitura de Curitiba. Os resultados poderão orientar o desenvolvimento de rotinas computacionais para pré-classificação dos protocolos, em apoio à tomada de decisão dos atendentes, e com a intenção de reduzir falhas humanas. Futuramente, respostas para determinados assuntos poderão ser emitidas automaticamente, diminuindo o tempo de espera dos cidadãos. Nessa perspectiva, o estudo almeja colaborar com a gestão da informação na prefeitura.

REFERÊNCIAS

ANATEL. Serviços de Utilidade Pública e de Emergência (SUP). Disponível em: <https://www.anatel.gov.br/setorregulado/codigos-nacionais/servicos-de-utilidade-publica-e-de-emergencia>. Acesso em: 28 jul. 2020.

ARANHA, C.; PASSOS, E. A Tecnologia de Mineração de Textos. **Revista Elerônica de Sistemas de Informação**, v. 5, n. 2, p. 1–8, 2006.

BRASIL. **Lei n. 13.460, de 26 de junho de 2017**. Dispõe sobre participação, proteção e defesa dos direitos do usuário dos serviços públicos da administração pública. Brasília, DF: Presidência da República, 2017.

CARRY COGLIANESE; LEHR, D. Transparency and Algorithmic Governance. **Administrative Law Review**, v. 71, n. 1, p. 18–38, 2019.

CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva Educação, 2016.

CURITIBA. **Lei n 11.095, de 21 de julho de 2004**. Dispõe sobre as normas que regulam a aprovação de projetos, o licenciamento de obras e atividades, a execução, manutenção e conservação de obras no Município, e dá outras providências. Curitiba: Câmara Municipal, 2004.

CURITIBA. **Portal de Dados Abertos da Prefeitura de Curitiba - 156**. 2019.

DAVENPORT, T. H.; PRUSAK, L. **Working Knowledge: How Organizations Manage What They Know**. Harvard Business Press, 1998.

DETLOR, B. Information management. **International Journal of Information Management**, v. 30, n. 2, p. 103–108, 2010.

FEINERER, I. *et al.* Text Mining Infrastructure in R. **Journal of Statistical Software**, v. 25, n. 5, 2008.

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO
02 a 04 de setembro de 2020

FELDMAN, R.; SANGER, J. **The text mining handbook: advanced approaches in analyzing unstructured data**. Cambridge university press, 2007.

KHANNA, R.; AWAD, M. **Efficient Learning Machines : Theories , Concepts , and Applications for Engineers and System Designers**. 2015.

LAROSE, D. T.; LAROSE, C. D. **Discovering knowledge in data: an introduction to data mining**. Second ed. Hoboken, New Jersey: Wiley, 2014.

LOPES, I. *et al.* **Inteligência Artificial**. 1^o ed. Rio de Janeiro: Elsevier, 2014.

MAIA, M. *et al.* Mineração de dados no apoio a gestão pública municipal: conhecendo as demandas da cidade de Curitiba pela “Central 156”. **R. Tecnol. Soc.**, v. 16, n. 40, p. 1–18, 2019. Disponível em: <https://periodicos.utfpr.edu.br/rts/article/view/8591>.

MARCHIORI, P. Z. A ciência e a gestão da informação: compatibilidades no espaço profissional. **Ciência da informação**, v. 31, n. 2, 2002.

MEHR, H. *et al.* **Artificial Intelligence for Citizen Services and Government**. Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., 2017.

MENDES JUNIOR; R. Crowdsourcing e machine learning: uma revisão sistemática com discussão do uso para a participação pública dos cidadãos. **Inteligência Artificial: 3o. Grupo de Pesquisa do ITS**, 2018. ITS Rio. Disponível em: <https://itsrio.org/wp-content/uploads/2019/03/Ricardo-Mendes.pdf>. Acesso em: 13 jul. 2020.

MENEZES, J. B. A participação popular como fonte de legitimidade democrática da administração pública. **X Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública**, p. 18–21, 2005. Chile.

MILLON, L. V. Participação popular nas políticas públicas. , v. 13, n. 17, p. 59–71, 2010. Disponível em: <https://revista.pgskroton.com/index.php/rdire/article/view/1896>.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. **Sistemas Inteligentes: fundamentos e aplicações**. p.89–114, 2003.

SCHNEIDER, C. The biggest data challenges that you might not even know you have. , p. 1–14, 2016. Disponível em: <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/>

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero Revista de Ciência da Informação**, v. 28, n. 0, p. 1–14, 1999.

SORDI, J. O. DE. **Administração da informação: fundamentos e práticas para uma nova gestão do conhecimento**. São Paulo: Saraiva Educação SA, 2008.

TAN, P.-N. *et al.* **Introduction to Data Mining**. Boston, USA: Pearson Addison-Wesley, 2006.

VASQUES, D. G. *et al.* Mineração de Textos para Gestão de Clientes em Empresas de

II SEMINÁRIO NACIONAL DE GESTÃO DA INFORMAÇÃO E CONHECIMENTO
02 a 04 de setembro de 2020

Telecomunicações. **XIII Brazilian Symposium on Information Systems**, p. 9–16, 2017.
Lavras, Minas Gerais.

WITTEN, I. H. *et al.* **Data mining practical machine learning tools and techniques**.
Cambridge, MA, USA: Morgan Kaufmann, 2017.