



DANIEL JOSÉ SILVA OLIVEIRA

**AVALIAÇÃO DO MÉTODO DE ANÁLISE DE
SENTIMENTO EM MÍDIAS SOCIAIS
APLICADO NA GESTÃO SOCIAL E POLÍTICA**

LAVRAS - MG

2015

DANIEL JOSÉ SILVA OLIVEIRA

**AVALIAÇÃO DO MÉTODO DE ANÁLISE DE SENTIMENTO EM
MÍDIAS SOCIAIS APLICADO NA GESTÃO SOCIAL E POLÍTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Administração Pública, área de concentração em Gestão de Organizações Públicas, para obtenção do título de Mestre.

Orientador

Prof. Paulo Henrique de Souza Bermejo, Dr.

LAVRAS - MG

2015

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Oliveira, Daniel José Silva.

Avaliação do método de análise de sentimento em mídias
sociais aplicado na Gestão Social e política / Daniel José Silva
Oliveira. – Lavras : UFLA, 2015.

111 p. : il.

Dissertação (mestrado)–Universidade Federal de Lavras, 2015.

Orientador(a): Paulo Henrique de Souza Bermejo.

Bibliografia.

1. Análise de sentimento. 2. Mídias sociais. 3. Administração
pública. 4. Gestão social. 5. Política. I. Universidade Federal de
Lavras. II. Título.

DANIEL JOSÉ SILVA OLIVEIRA

**AVALIAÇÃO DO MÉTODO DE ANÁLISE DE SENTIMENTO EM
MÍDIAS SOCIAIS APLICADO NA GESTÃO SOCIAL E POLÍTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Administração Pública, área de concentração em Gestão de Organizações Públicas, para obtenção do título de Mestre.

APROVADA em 30 de janeiro de 2015.

Dr. José Roberto Pereira UFLA
Dra. Rita de Cássia Romeiro Paulino UFSC

Prof. Paulo Henrique de Souza Bermejo, Dr.
Orientador

LAVRAS - MG

2015

*Às três mulheres que me inspiram, motivam e
me fazem querer ser uma pessoa melhor a cada dia.
Minha avó Edite (in memoriam), minha mãe Vanilda e minha noiva Daniely.*

DEDICO

AGRADECIMENTOS

Agradeço primeiramente a Deus por guiar meus passos e me dar forças para superar obstáculos.

À minha mãe Vanilda que acompanhou de perto meus estudos e me deu todo o apoio necessário, sempre me fazendo acreditar em minha capacidade de conquistar objetivos.

À minha noiva Daniely pelo carinho, dedicação, por me ajudar nos momentos difíceis e pela cumplicidade.

Ao meu orientador, professor Paulo Henrique de Souza Bermejo, pelo incentivo e dedicação. Sobretudo, agradeço-lhe pela confiança, paciência e por todos os ensinamentos.

À coordenadora do Programa de Pós - Graduação em Administração Pública, professora Sílvia Helena Rigatto, e demais professores do programa pelos ensinamentos e contribuições para a conclusão deste trabalho, e às secretárias Nádia e Andréia pelo apoio em todas as fases do curso.

Aos colegas de curso, pela amizade, pelo companheirismo e pelas contribuições tanto nas disciplinas quanto neste trabalho.

Ao Dr. Stuart Shulman, professor da *University of Massachusetts Amherst* e criador do aplicativo DiscoverText, por conceder uma licença acadêmica do referido aplicativo, que foi substancial para a conclusão deste trabalho.

Aos funcionários da UFLA, em especial, àqueles ligados diretamente ao Departamento de Administração e Economia.

Enfim, a todos que, de maneira direta ou indireta, contribuíram para a realização desta dissertação, que é resultado de muito trabalho e longas estradas.

“Acima de todas as liberdades, dê-me a de saber, de me expressar, de debater com autonomia, de acordo com minha consciência.”

John Milton

“A emergência de um público ávido por discutir questões de interesse comum modificou a estrutura sob a qual a sociedade se alicerçava.”

Jürgen Habermas

RESUMO

Este estudo é uma dissertação de mestrado apresentada sob a forma de três artigos de natureza teórico-empírica. O objetivo geral deste estudo é avaliar a análise de sentimento, realizada a partir de dados extraídos de mídias sociais, como um método que pode ser utilizado para mensurar a opinião pública sobre temas relacionados à política e à administração pública, gerando conhecimentos que possibilitem incorporar a opinião da sociedade civil nas decisões políticas do Estado. Para atingir esse objetivo, os três artigos abordaram as principais formulações teóricas e conceituais acerca do tema e apresentaram demonstrações práticas, por meio da aplicação da técnica de análise de sentimento, em dados extraídos de mídias sociais, relacionados à política e à administração pública. Trata-se de um estudo aplicado, interdisciplinar e exploratório que pode ser classificado como qualitativo e quantitativo. O primeiro artigo buscou avaliar se a análise de sentimento pode ser utilizada como um instrumento de mensuração da opinião pública sobre a atuação do governo. Para fins de demonstração, a análise de sentimento foi aplicada em dados extraídos do Facebook e Twitter, a partir de temas relacionados aos fatores que motivaram a onda de protestos realizados no Brasil em junho de 2013. Por meio desta análise, foi observado que a análise de sentimento pode revelar a opinião dos cidadãos sobre temas relacionados à política e à administração pública. O segundo artigo descreve como a análise de sentimento pode contribuir para as práticas da Gestão Social colaborando com a consolidação de um modelo de gestão pública mais democrático. O estudo consistiu na pesquisa empírica e no confronto entre os conceitos e aplicações, envolvendo Gestão Social, análise de sentimento e as estratégias de utilização de mídias sociais adotadas pelos governos. Ao final, o estudo demonstrou que a análise de sentimento pode contribuir para as práticas da Gestão Social no contexto da estratégia de rede. O terceiro artigo procurou identificar se a análise de sentimento pode revelar precisamente as preferências políticas dos cidadãos se comparada a pesquisas de opinião tradicionais. Para tanto, foi realizada a análise de sentimento, a partir de *tweets* relacionados aos candidatos a presidente da república, durante o segundo turno das eleições presidenciais no Brasil, em 2014. Posteriormente, os resultados dessa análise foram confrontados com os resultados de seis pesquisas de intenção de votos e de rejeição dos candidatos realizada pelo Instituto de Pesquisas Datafolha. Os resultados deste estudo demonstraram que a análise de sentimento pode indicar a preferência dos eleitores com resultados aproximados das pesquisas de opinião tradicionais.

Palavras-chave: Análise de sentimento. Mídias sociais. Administração pública. Gestão social. Política.

ABSTRACT

This study is a Master's degree thesis presented in the form of three articles of theoretical-empirical nature. The general objective of this study is to evaluate the sentiment analysis, performed based on data extracted from social media, as a method that can be used to measure public opinion on themes related to politics and public administration, generating knowledge that allow us to incorporate the opinion of civilian society on the political decisions of the State. To reach this objective, the three articles approached the main theoretical and conceptual formulations regarding the theme and presented practical demonstrations by means of applying the technique of sentiment analysis on data extracted from social media related to politics and public administration. This is an applied, interdisciplinary and exploratory study that can be classified as qualitative and quantitative. The first article sought to evaluate if the sentiment analysis can be used as an instrument to measure public opinion on Government actions. For demonstrative purposes, the sentiment analysis was applied to data extracted from Facebook and Twitter, based on themes related to factors that motivate the wave of protestations occurring in Brazil in June of 2013. With this analysis, we observed that the sentiment analysis could reveal the opinion of citizens on themes related to politics and public administration. The second article describes how the sentiment analysis can contribute on Social Management practices, collaborating with the consolidation of a more democratic public management model. The study consisted on the empirical research and on the confrontation between concepts and applications, involving Social Management, sentiment analysis and the strategies of social media use adopted by the governments. Finally, the study showed that sentiment analysis could contribute with Social Management practices in the context of network strategy. The third article sought to identify if the sentiment analysis can precisely reveal the political preferences of the citizens if compared to traditional opinion survey. For this, we conducted a sentiment analysis based on tweets regarding the candidates for Republic Presidency, for the second round of the presidential elections in Brazil in 2014. Posteriorly, the results of this analysis were confronted with the results of six voting intention and rejecting candidates polls performed by the Instituto de Pesquisas Datafolha. He results of this study showed that the sentiment analysis could indicate the preference of electors with results approximating the traditional opinion polls.

Keywords: Sentiment analysis. Social media. Public administration. Social Management. Politics.

LISTA DE GRÁFICOS

ARTIGO 1

Gráfico 1	Análise de sentimento dos brasileiros baseada nos dados coletados em dezembro de 2013 e junho de 2014.....	41
-----------	--	----

ARTIGO 2

Gráfico 1	Análise de sentimento dos brasileiros em relação a quatro programas sociais.....	68
-----------	--	----

ARTIGO 3

Gráfico 1	Resultados da análise de sentimento aplicada em seis conjuntos de dados contendo <i>tweets</i> relacionados aos candidatos a presidente da república nas eleições presidenciais de 2014 no Brasil.....	96
Gráfico 2	Resultados das seis pesquisas de intenção de votos para presidente da república no segundo turno.....	98
Gráfico 3	<i>Tweets</i> positivos relacionados aos candidatos.....	99
Gráfico 4	Comparação entre a pesquisa Datafolha (2014) de intenção de votos e os <i>tweets</i> positivos identificados na análise de sentimento.....	99
Gráfico 5	Resultados das seis pesquisas de rejeição dos candidatos a presidente da república.....	100
Gráfico 6	<i>Tweets</i> negativos relacionados aos candidatos.....	101
Gráfico 7	Comparação entre a pesquisa Datafolha (2014) de rejeição dos candidatos e os <i>tweets</i> negativos identificados na análise de sentimento.....	101

LISTA DE QUADROS

ARTIGO 1

Quadro 1	Termos de pesquisa utilizados na mineração de dados	36
----------	---	----

ARTIGO 2

Quadro 1	Palavras-chave utilizadas na mineração de dados	64
----------	---	----

Quadro 2	Exemplo de <i>tweets</i> classificados pela polaridade.....	69
----------	---	----

Quadro 3	Relação entre estratégias de utilização de mídias sociais, Gestão Social e análise de sentimento.....	69
----------	--	----

ARTIGO 3

Quadro 1	Termos-chave utilizados na mineração de dados	92
----------	---	----

Quadro 2	Exemplo de <i>tweets</i> classificados pela polaridade.....	97
----------	---	----

LISTA DE TABELAS

ARTIGO 1

Tabela 1 Cálculo do conjunto de treino 38

Tabela 2 Precisão da análise de sentimento 40

ARTIGO 2

Tabela 1 Cálculo do conjunto de treino 65

Tabela 2 Precisão da análise automatizada dos dados 66

ARTIGO 3

Tabela 1 Distribuição dos conjuntos de dados 93

Tabela 2 Conjunto de treino 94

Tabela 3 Precisão da classificação automatizada dos dados 95

SUMÁRIO

	PRIMEIRA PARTE	14
1	INTRODUÇÃO	14
	REFERÊNCIAS	19
	SEGUNDA PARTE – ARTIGOS	22
	ARTIGO 1: Análise de Sentimento, Mídias Sociais e Administração Pública	22
1	INTRODUÇÃO	23
2	A RELAÇÃO ENTRE CIDADÃOS E GOVERNOS NAS MÍDIAS SOCIAIS	25
3	OPINIÃO PÚBLICA E MÍDIAS SOCIAIS	27
4	ANÁLISE DE SENTIMENTO	30
4.1	Vantagens e limitações da técnica	32
5	TRABALHOS RELACIONADOS	33
6	A ANÁLISE DE SENTIMENTO NA PRÁTICA	34
6.1	Procedimentos Metodológicos	35
7	RESULTADOS E DISCUSSÃO	40
8	TRABALHOS FUTUROS	43
9	CONCLUSÃO	44
	REFERÊNCIAS	46
	ARTIGO 2: A análise de sentimento em mídias sociais como instrumento para as práticas da Gestão Social	52
1	INTRODUÇÃO	53
2	ESTRATÉGIAS DE UTILIZAÇÃO DE MÍDIAS SOCIAIS PELA ADMINISTRAÇÃO PÚBLICA	54
3	PARTICIPAÇÃO SOCIAL E A PROMOÇÃO DA DEMOCRACIA	56

4	GESTÃO SOCIAL	57
5	ANÁLISE DE SENTIMENTO	60
6	METODOLOGIA	62
7	RESULTADOS E DISCUSSÃO	67
8	CONCLUSÃO	71
	REFERÊNCIAS	73
	ARTIGO 3: As mídias sociais podem revelar as preferências dos eleitores? Um confronto entre análise de sentimento e pesquisas de opinião tradicionais	79
1	INTRODUÇÃO	80
2	PESQUISAS DE OPINIÃO EM MÍDIAS SOCIAIS	82
3	ANÁLISE DE SENTIMENTO EM MÍDIAS SOCIAIS	84
4	TWITTER COMO FONTE DE DADOS	86
5	TRABALHOS RELACIONADOS	88
6	METODOLOGIA	90
7	RESULTADOS E DISCUSSÃO	96
8	CONCLUSÃO	102
	REFERÊNCIAS	104
	CONSIDERAÇÕES FINAIS	109
	ANEXO	111

PRIMEIRA PARTE

1 INTRODUÇÃO

As interações facilitadas pelas mídias sociais tornaram-se parte integrante da vida cotidiana das pessoas na sociedade contemporânea, transformando, significativamente, a maneira como estas se comunicam (CRIADO, SANDOVAL-ALMAZAN; GIL-GARCIA, 2013; KONTOPOULOS et al., 2013). As várias plataformas de mídia social como *blogs*, *microblogs*, *wikis* e redes sociais eletrônicas facilitam a interação entre usuários proporcionando-lhes a oportunidade de compartilhar informações e opiniões sobre vários temas, que podem envolver desde assuntos triviais (marcas, produtos, etc.) até questões de interesse público (SOBKOWICZ, KASCHEKY; BOUCHARD, 2012; ZHANG; HE, 2013; KHAN et al., 2014).

O grande volume de opiniões postadas nessas mídias tem influenciado o comportamento social, político e econômico em todo o mundo (MONTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012). Nos campos da política e da administração pública, por exemplo, as opiniões expressas nas mídias sociais podem influenciar desde os candidatos políticos que os cidadãos apoiam até o comportamento da opinião pública perante a atuação do governo (EIRINAKI; PISAL; SINGH, 2012; MOSTAFA, 2013).

Principalmente com a popularização do acesso à internet por meio de diversas tecnologias, tais como computadores, *smartphones*, *tablets*, e outros dispositivos conectados à web, a utilização das mídias sociais tem crescido em ritmo acelerado (CERON et al., 2014). Consequentemente, uma grande quantidade de conteúdo gerado pelos usuários dessas mídias está disponível e acessível, o que representa uma valiosa fonte de dados para identificar a opinião das pessoas sobre diversos assuntos (CERON et al., 2014; HE; ZHA; LI, 2013).

Mas o desafio é transformar essa grande quantidade de dados não estruturados em conhecimento (LIU, 2010; MONTOYO, MARTÍNEZ-BARCO; BALAHUR, 2012; MERGEL, 2013). Na literatura, entre as tecnologias que possibilitam gerar conhecimento baseado na coleta, processamento e análise de dados extraídos das mídias sociais, destaca-se a análise de sentimento, também conhecida como mineração de opinião (LIU; ZHANG, 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012; ZHANG; HE, 2013).

A análise de sentimento é uma técnica automatizada de descoberta de conhecimento, a partir da mineração de dados textuais que tem o objetivo de identificar a opinião das pessoas sobre temas específicos (DI CARO; GRELLA, 2013; MOSTAFA, 2013). Seu propósito é criar uma base de conhecimento contendo opiniões polarizadas (positivas, negativas ou neutras), por meio de ferramentas que processam automaticamente grandes volumes de dados não estruturados, cuja leitura e análise humana seriam inviáveis (LIU; ZHANG, 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

Nesse contexto, a presente dissertação de mestrado foi realizada com o objetivo geral de avaliar a análise de sentimento, realizada a partir de dados extraídos de mídias sociais, como um método que pode ser utilizado para mensurar a opinião pública sobre temas relacionados à política e à administração pública, gerando conhecimentos que possibilitem incorporar a opinião da sociedade civil nas decisões políticas do Estado.

Esta dissertação é apresentada sob a forma de três artigos de natureza teórico-empírica, que conjuntamente materializam seus objetivos específicos, que são:

- a) avaliar a análise de sentimento como um instrumento de mensuração da opinião pública sobre a atuação do governo, contribuindo para que o mesmo possa atuar em consonância com o interesse público;

- b) investigar como a técnica de análise de sentimento pode contribuir para as práticas da Gestão Social; e
- c) confrontar a análise de sentimento com pesquisas de opinião tradicionais sobre as preferências políticas dos cidadãos.

Para atingir estes objetivos, os três artigos abordaram as principais formulações teóricas e conceituais acerca do tema e apresentaram demonstrações práticas aplicando-se a técnica de análise de sentimento em dados extraídos de mídias sociais relacionados à política e à administração pública, utilizando-se o aplicativo DiscoverText¹.

O primeiro artigo intitulado “Análise de Sentimento, Mídias Sociais e Administração Pública” buscou identificar se a análise de sentimento pode ser utilizada como um instrumento de mensuração da opinião pública sobre a atuação do governo de forma a contribuir para a avaliação da administração pública, no sentido de alcançar o interesse público e, assim, evitar impopularidade ou mesmo conflitos. Para fins de demonstração, foram selecionados temas que motivaram a realização das manifestações populares de junho de 2013 no Brasil. Esses temas foram selecionados com base em um levantamento sobre os fatores que motivaram tais manifestações, realizado pelo Instituto Brasileiro de Opinião Pública e Estatística (IBOPE, 2013). Foram coletadas, processadas e analisadas, aproximadamente, 130.000 mensagens postadas no Facebook e no Twitter considerando esses temas. Por meio desta análise, pôde-se observar que, mesmo após um ano, os fatores que motivaram as manifestações continuam gerando insatisfação nos cidadãos. Diante disso, a

¹ O DiscoverText (<http://www.discovertext.com>) é um aplicativo de análise de textos, baseado em nuvem, capaz de capturar, filtrar e classificar um grande volume de dados não estruturados, que além de facilitar a captura de dados de mídia social, fornece ferramentas analíticas bem integradas para explorar e codificar os dados (BEYER, 2012).

análise de sentimento pode contribuir no monitoramento da opinião dos cidadãos quanto à atuação do governo e, também, apontar, por meio da frequência de opiniões positivas e negativas, o quanto eles estão satisfeitos ou insatisfeitos com os serviços públicos.

O segundo artigo intitulado “A análise de sentimento como instrumento para as práticas da Gestão Social” procurou demonstrar como a análise de sentimento pode ser utilizada para estabelecer um fluxo de comunicação entre Estado e sociedade contribuindo para incorporar a opinião da sociedade civil na administração pública. A aplicação da análise de sentimento foi demonstrada por meio de dados extraídos do Twitter sobre alguns dos principais programas sociais implementados no Brasil pelo Governo Federal nos últimos anos. Ao confrontar os resultados observados nesse exemplo com os conceitos presentes na literatura sobre Gestão Social, análise de sentimento e as estratégias de utilização de mídias sociais adotadas pelos governos, foi possível perceber que a análise de sentimento pode contribuir para as práticas da Gestão Social no contexto da estratégia de rede.

O terceiro artigo intitulado “As mídias sociais podem revelar as preferências dos eleitores? Um confronto entre análise de sentimento e pesquisas de opinião tradicionais” procurou identificar se os resultados, obtidos pela aplicação da técnica de análise de sentimento em dados extraídos do Twitter podem revelar as preferências políticas dos cidadãos com precisão aproximada das pesquisas de opinião tradicionais. Para tanto, foram coletados, preparados e analisados 92.441 *tweets*, a partir de termos-chave relacionados aos candidatos Aécio Neves e Dilma Rousseff, durante o segundo turno das eleições presidenciais do Brasil, em 2014, revelando a opinião positiva, negativa e neutra dos cidadãos perante estes candidatos. Posteriormente, os resultados dessa análise foram confrontados com os resultados de seis pesquisas de intenção de votos e de rejeição dos candidatos realizada pelo Instituto de Pesquisa Datafolha.

Os resultados deste estudo demonstraram que a análise de sentimento pode indicar a preferência dos eleitores com resultados aproximados aos das pesquisas de opinião tradicionais.

A presente dissertação se justifica porque, atualmente, as mídias sociais se tornaram importantes canais de interação que podem fazer a diferença na opinião dos cidadãos em relação ao governo (MERGEL, 2013). E, mesmo com o número crescente de pesquisas que utilizam essas mídias como fonte de dados, “os pesquisadores têm pouco conhecimento sobre como aplicar os métodos de mineração de conteúdo” (YOON; ELHADAD; BAKKEN, 2013, p. 122). Este trabalho tem o propósito de preencher algumas lacunas identificadas na literatura, como: avaliar práticas avançadas de uso de mídias sociais no contexto da administração pública em que as opiniões expressas nessas mídias possam ser organizadas gerando conhecimento (FERRO et al., 2013); verificar se as mídias sociais possuem potencial para aumentar a participação social e fortalecer a democracia estabelecendo uma relação dialógica entre Estado e sociedade (SNEAD, 2013; SOBACI; KARKIN, 2013; OLIVEIRA; BERMEJO; SANTOS, 2015); e testar a eficácia da análise automatizada de grandes volumes de dados não estruturados confrontando-os com métodos de pesquisa tradicionais (VARGO et al., 2014).

Este estudo também traz contribuições teóricas e empíricas relevantes no que diz respeito às etapas a serem percorridas para identificar as opiniões expressas nas mídias sociais, por meio da análise de sentimento, bem como reconhece algumas limitações da técnica e aponta direções para estudos futuros.

REFERÊNCIAS

BEYER, Y. Using discovertext for large scale twitter harvesting. **Microform & Digitization Review**, London, v. 41, n. 3-4, p. 121-125, Dec. 2012.

CERON, A. et al. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. **New Media & Society**, Milano, v. 16, n. 2, p. 340-358, Mar. 2014.

CRIADO, J. I.; SANDOVAL-ALMAZAN, R.; GIL-GARCIA, J. R. Government innovation through social media. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 319-326, Oct. 2013.

DI CARO, L.; GRELLA, M. Sentiment analysis via dependency parsing. **Computer Standards & Interfaces**, Amsterdam, v. 35, n. 5, p. 442-453, Sept. 2013.

EIRINAKI, M.; PISAL, S.; SINGH, J. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, New York, v. 78, n. 4, p. 1175-1184, July 2012.

FERRO, E. et al. Policy making 2.0: from theory to practice. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 359-368, Oct. 2013.

HE, W.; ZHA, S.; LI, L. Social media competitive analysis and text mining: a case study in the pizza industry. **International Journal of Information Management**, Guildford, v. 33, n. 3, p. 464-472, June 2013.

IBOPE. 89% dos manifestantes não se sentem representados por partidos. **Instituto Brasileiro de Opinião Pública e Estatística**, Rio de Janeiro, abr. 2013. Disponível em: <<http://www.ibope.com.br/pt-br/noticias/Paginas/89-dos-manifestantes-nao-se-sentem-representados-por-partidos.aspx>>. Acesso em: 11 out. 2013

KHAN, G. F. et al. From e-government to social government: twitter use by Korea's central government. **Online Information Review**, Bradford, v. 38, n. 1, p. 95-113, Jan. 2014.

KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4065-4074, Aug. 2013.

LIU, B. **Sentiment analysis and subjectivity**. 2. ed. Chicago: Handbook of Natural Language Processing, 2010.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining text data**. Chicago: Springer, 2012. Cap. 13, p. 415-463.

MERGEL, I. A framework for interpreting social media interactions in the public sector. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 327-334, Oct. 2013.

MONTOYO, A.; MARTÍNEZ-BARCO, P.; BALAHUR, A. Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. **Decision Support Systems**, Amsterdam, v. 53, n. 4, p. 675-679, Nov. 2012.

MOSTAFA, M. M. More than words: social networks' text mining for consumer brand sentiments. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4241-4251, Aug. 2013.

OLIVEIRA, D. J. S.; BERMEJO, P. H. S.; SANTOS, P. A. Sentiment analysis, social media, and public administration. In: CEMAL, D. et al. (Ed.). **Handbook of research on democratic strategies and citizen-centered E-government services**. Hershey: IGI Global, 2015. Cap. 13, p. 231-249.

SNEAD, J. T. Social media use in the U.S. executive branch. **Government Information Quarterly**, Amsterdam, v. 30, n. 1, p. 56-63, Jan. 2013.

SOBACI, M. Z.; KARKIN, N. The use of twitter by mayors in Turkey: tweets for better public services? **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 417-425, Oct. 2013.

SOBKOWICZ, P.; KASCHEKY, M.; BOUCHARD, G. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 470-479, Oct. 2012.

VARGO, C. J. et al. Network issue agendas on twitter during the 2012 U.S. presidential election. **Journal of Communication**, New York, v. 64, n. 2, p. 296-316, Mar. 2014.

YOON, S.; ELHADAD, N.; BAKKEN, S. A practical approach for content mining of tweets. **American Journal of Preventive Medicine**, New York, v. 45, n. 1, p. 122-129, July 2013.

ZHANG, P.; HE, Z. A weakly supervised approach to Chinese sentiment classification using partitioned self-training. **Journal of Information Science**, Cambridge, v. 39, n. 6, p. 815-831, Dec. 2013.

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

Análise de Sentimento, Mídias Sociais e Administração Pública²

RESUMO

Este estudo busca identificar como a análise de sentimento, baseada em textos extraídos de mídias sociais, pode ser um instrumento de mensuração da opinião pública sobre a atuação do governo de forma a contribuir para a avaliação da administração pública. Trata-se de um estudo aplicado, interdisciplinar, exploratório, qualitativo e quantitativo. Foram revisadas as principais formulações teóricas e conceituais acerca do tema e realizadas demonstrações práticas, utilizando-se uma ferramenta de mineração de opinião que proporcionou precisão satisfatória no processamento de dados. Para fins de demonstração, foram selecionados temas que motivaram a realização das manifestações populares que envolveram milhões de pessoas no Brasil em junho de 2013. Foram coletadas, processadas e analisadas, aproximadamente, 130.000 mensagens postadas no Facebook e no Twitter, sobre esses temas em dois períodos distintos. Por meio desta análise, observou-se que a análise de sentimento pode revelar a opinião dos cidadãos quanto à atuação do governo.

Palavras-chave: Análise de sentimento. Mineração de opinião. Mídias sociais. Administração pública.

² Este estudo foi publicado em coautoria com Paulo Henrique de Souza Bermejo e Pâmela Aparecida dos Santos em sua versão na língua inglesa como um dos capítulos do livro *Handbook of Research on Democratic Strategies and Citizen-Centered E-Government Services*, organizado por Čemal Dolićanin, Ejub Kajan, Dragan Randjelović e Boban Stojanović, e publicado pela editora IGI-Global, Hershey, PA - USA, em 13 de janeiro de 2015.

1 INTRODUÇÃO

O grande volume de dados postados na internet, por meio das mídias sociais, está produzindo mudanças importantes na maneira pela qual as pessoas se comunicam, compartilham conhecimentos e emoções que influenciam o comportamento social, político e econômico em todo o mundo (MONTYO; MARTÍNEZ-BARCO; BALAHUR, 2012).

De acordo com Mostafa (2013), as opiniões expressas nas mídias sociais desempenham um papel importante em influenciar o comportamento da opinião pública em diversas áreas. A título de exemplo, nas áreas de política e administração pública, essas mídias têm o poder de disseminar opiniões que podem resultar tanto em melhorias nos serviços públicos quanto em protestos motivados pela insatisfação dos cidadãos com os governos (PAPACHARISSI; OLIVEIRA, 2012; O'CALLAGHAN et al., 2014).

A popularidade dessas mídias é considerável, tornando-as uma significativa fonte de informações que, também, pode ser utilizada na melhoria dos serviços públicos (MERGEL, 2013). Portanto, além dos canais diretos em que o cidadão pode emitir sua opinião como e-mail, ouvidorias e portais de reclamação, as mídias sociais (*blogs, microblogs, wikis, fóruns, redes sociais eletrônicas, entre outras*) podem ser utilizadas para promover serviços públicos participativos e orientados para o cidadão (SOBACI; KARKIN, 2013). Mas o grande volume de informações que circula atualmente nessas plataformas demanda tecnologias que permitam sua análise (BONSÓN et al., 2012). Uma opção que vem ganhando espaço entre os pesquisadores é a análise de sentimento (PANG; LEE, 2008; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

A análise de sentimento ou mineração de opinião pode ser considerada uma técnica automatizada de descoberta de conhecimento que visa encontrar padrões ocultos em um grande número de informações textuais, inclusive,

mídias sociais (MOSTAFA, 2013). Seu objetivo é criar uma base de conhecimento contendo opiniões (positivas, negativas e neutras) de uma forma mais estruturada e explícita que expressem sentimentos, avaliações e percepções das pessoas em relação a qualquer tema (FORTUNY et al., 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

Neste contexto, o presente estudo foi realizado com o objetivo de avaliar se a análise de sentimento pode ser utilizada como um instrumento de mensuração da opinião pública sobre a atuação do governo, contribuindo para que o mesmo possa atuar em consonância com o interesse público, evitando impopularidade ou mesmo conflitos. Assim, a questão que norteia este estudo é: A análise de sentimento em mídias sociais pode mensurar a opinião dos cidadãos sobre a atuação do governo e dessa forma contribuir para a melhoria dos serviços públicos?

Para responder essa questão, esta pesquisa explorará a técnica de análise de sentimento, baseada na mineração de dados em mídias sociais sobre temas de grande relevância para a administração pública demonstrando sua capacidade para identificar a opinião dos cidadãos. Pretende-se, assim, preencher uma lacuna identificada na literatura por Ferro et al. (2013), que destacam a necessidade de se avaliar práticas avançadas de uso de mídias sociais no contexto da administração pública, em que as opiniões expressas nessas mídias possam ser organizadas gerando conhecimento.

Este trabalho se justifica porque é necessário entender o uso de aplicativos de mídia social como canais aceitáveis para as interações entre governo e seus diversos públicos, e que podem, potencialmente, fazer a diferença nas percepções e sentimentos dos cidadãos em relação ao governo (MERGEL, 2013). E, apesar da crescente atenção à análise de conteúdo gerado pelo usuário, em mídias sociais, “os pesquisadores têm pouco conhecimento sobre como aplicar os métodos de mineração de conteúdo” (YOON; ELHADAD;

BAKKEN, 2013, p. 122), faltando métricas adequadas que identifiquem, nas mídias sociais, os impactos das ações dos governos (MERGEL, 2013; HOFMANN et al., 2013).

Este estudo está dividido em nove seções. Na próxima seção é discutida a relação entre cidadãos e governos nas mídias sociais. Na terceira seção, é abordada a correlação entre as mídias sociais e a opinião pública. Na quarta seção, é explorada a literatura recente sobre análise de sentimento, descrevendo as principais vantagens e limitações da técnica. Na quinta seção, são apresentados alguns trabalhos relacionados ao tema. Na sexta seção, são demonstrados os procedimentos metodológicos para a aplicação da análise de sentimento por meio de um exemplo prático. Os resultados são apresentados na sétima seção. Na oitava seção, são apontados alguns direcionamentos para estudos futuros. Finalmente, na nona seção, o estudo é concluído.

2 A RELAÇÃO CIDADÃO-GOVERNO NAS MÍDIAS SOCIAIS

Os cidadãos estão cada vez mais utilizando as mídias sociais para se comunicarem com seus familiares, amigos, colegas de trabalho, empresas e até mesmo com o governo (KAVANAUGH et al., 2012). Essas mídias compreendem um conjunto de tecnologias que permitem a comunicação entre os cidadãos e o governo de forma colaborativa (OLIVEIRA; WELCH, 2013). Exemplo disso é o número crescente de usuários de mídias sociais interessados em vários assuntos relacionados à política e à administração pública (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012). Eles cobrem todo o espectro de interessados, desde simples cidadãos que expressam as suas opiniões sobre as questões do cotidiano, a políticos que utilizam este meio para comunicar suas ideias, além de jornalistas que criticam o governo ou o próprio governo como forma de divulgar suas ações (TSYTSARAU; PALPANAS, 2012).

Até mesmo em regimes antidemocráticos o habitante do ciberespaço reúne graus de liberdade para manifestar sua opinião fazendo com que as mídias sociais se tornem ambientes propícios para a difusão de opiniões favoráveis ou contrárias às ações dos governos (CÁCERES, 2011). Estudos recentes demonstram que as mídias sociais foram utilizadas por cidadãos de vários países na articulação de manifestações contra seus governos, como a Primavera Árabe que foi marcada por protestos e conflitos no Oriente Médio e no norte da África destacando-se os conflitos ocorridos no Egito (CHEN, 2011; GLASS; COLBAUGH, 2012; LEWIŃSKI; MOHAMMED, 2012; LIM, 2012; PAPACHARISSI; OLIVEIRA, 2012) e na Síria (O'CALLAGHAN et al., 2014). Em junho de 2013, uma série de protestos, também, foram organizados no Brasil por meio das mídias sociais. Segundo dados de uma pesquisa, realizada pelo Instituto Brasileiro de Opinião Pública e Estatística (IBOPE, 2013a) em um dos protestos de maior vulto, 62% dos manifestantes souberam do evento pelo Facebook, e 75% destes convocaram outras pessoas para participar do movimento por meio das plataformas Facebook e Twitter.

Muitas organizações privadas, principalmente as grandes corporações, já perceberam a importância da opinião expressa por seus clientes nas mídias sociais (CAMBRIA et al., 2013). Por outro lado, a administração pública é um dos campos que são mais claramente distantes da dinâmica do ciberespaço (CÁCERES, 2011). Para Hofmann et al. (2013), os benefícios a serem obtidos com a utilização das informações disponíveis nas mídias sociais não estão limitadas às organizações do setor privado, mas devem, também, estender-se aos governos. De acordo com Sobkowicz, Kaschesky e Bouchard (2012), até recentemente os formuladores de políticas públicas não tinham muitos indicadores disponíveis sobre a opinião dos cidadãos, e o que a maioria das pessoas sentia e pensava sobre as ações dos governos era algo inacessível. Atualmente, alguns governos já utilizam as mídias sociais com o objetivo de

promover a transparência, a participação cidadã e a colaboração, além de poderem contar com mais uma fonte de informação que contribui no processo decisório (CHUN; LUNA-REYES, 2012; MERGEL, 2013; MOSSBERGER; WU; CRAWFORD, 2013; SNEAD, 2013).

Para entender como as mídias sociais podem, dessa forma, serem utilizadas para estreitar a relação entre os cidadãos e os governos, é preciso conhecer seu potencial como uma fonte de dados capaz de refletir a opinião pública.

3 OPINIÃO PÚBLICA E MÍDIAS SOCIAIS

A opinião pública é formada por um processo de expressão coletiva de opiniões e tomada de decisão, dentro de uma esfera pública vibrante, em que os cidadãos podem trocar informações e ideias sobre questões públicas (PERRIN; MCFARLAND, 2011; STEENKAMP; HYDE-CLARKE, 2014). De acordo com Habermas (2003), a esfera pública se reproduz por meio do agir comunicativo, implicando apenas o domínio de uma linguagem natural que esteja em sintonia com a prática comunicativa cotidiana. Nesse contexto, Steenkamp e Hyde-Clarke (2014, p. 92) acreditam que as mídias sociais podem atuar como instituições comunicativas que representam uma expansão do espaço no qual as expressões de unidade ou diversidade possam ser comunicadas em uma “esfera pública interativa”.

As mídias sociais são aplicativos, baseados na internet, projetados para promover a interação social, a colaboração, a aprendizagem conjunta e a rápida difusão de informações pela sociedade (BONSÓN et al., 2012; KAVANAUGH et al., 2012). Indiscutivelmente, essas mídias tornaram-se plataformas de comunicação bastante populares onde os cidadãos podem ter acesso às opiniões de outras pessoas e expressarem sua própria opinião sobre qualquer tema

(ZHANG et al., 2009). Dessa forma, cidadãos investidos no bem comum podem usar espaços de discussão on-line para produzir uma opinião pública informada que pode avaliar e até guiar processos oficiais de tomada de decisão (PENNEY, 2014). Isso corrobora com o pensamento de Habermas (2003), em que a opinião pública representa o poder de influência da sociedade, para direcionar ou mesmo interferir nas decisões de parlamentares, governos e tribunais, sem tomar para si o poder administrativo.

A esfera pública interativa, criada por meio das mídias sociais, permite que os cidadãos se mantenham atualizados sobre qualquer assunto político ou social, no bairro, no estado, no país e em todo o mundo (EIRINAKI; PISAL; SINGH, 2012; STEENKAMP; HYDE-CLARKE, 2014). O fato de que a internet está agora ao alcance da maioria das pessoas em razão dos *hardwares* com preços mais acessíveis, levou a uma revolução cultural, pois pessoas de todo o mundo agora são capazes de interagir uns com os outros manifestando suas opiniões livremente (MOREO et al., 2012).

As mídias sociais estão mudando a forma com que as pessoas buscam informações e se comunicam (OLIVEIRA; WELCH, 2013). Percebe-se que o panorama da mídia mudou muito nos últimos anos, pois antes predominavam as mídias tradicionais unidirecionais (por exemplo, jornais, revistas e TV) e agora elas estão sendo complementadas ou substituídas por meios interativos de comunicação social (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012). De acordo com Fortuny et al. (2012), em um cenário imperfeito, mídias tendenciosas poderiam editar e selecionar informações com o objetivo de influenciar a opinião pública. Isso acontece, quando há a formação de monopólios nos sistemas de mídias tradicionais, que impedem a manifestação das esferas comunicativas plurais, o que, conseqüentemente, compromete a qualidade da representação das vozes públicas (GUIMARÃES; AMORIM, 2013). Por isso, “a formação democrática da opinião e da vontade depende de opiniões públicas

informais que idealmente se formam em estruturas de uma esfera pública não desvirtuada pelo poder” (HABERMAS, 2003, p. 33). Nesse sentido, o conhecimento obtido com base em mídias sociais é extremamente valioso, porque milhões de opiniões são expressas sobre diversos temas, sendo improvável que estas informações sejam tendenciosas (MOSTAFA, 2013). E mesmo se alguém, por meio das mídias sociais, tente influenciar a opinião pública em uma direção, haverá, também, aqueles que tentarão influenciá-la para outra direção e, assim, o volume de mensagens concorrentes simultâneas acaba anulando-se mutuamente (BACHNER; HILL, 2014). Conforme descreve Nunomura (2013), se por um lado, ainda, é preciso mais estudos para conferir às mídias sociais o status de um termômetro da opinião pública, por outro já se percebe que elas amplificam esse conceito.

Assim, as mídias sociais tornaram-se fontes de informação capazes de refletir a opinião pública sobre qualquer assunto, inclusive, política e administração pública (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012). Conseqüentemente, nos últimos anos têm aumentado o número de pesquisas acadêmicas e corporativas que exploram as informações contidas nessas mídias (YU; DUAN; CAO, 2013). Em contraste com os tempos mais antigos, quando encontrar fontes de informação era o principal problema para pesquisadores, organizações e indivíduos, a sociedade da informação de hoje desafia os mesmos a criarem e implementarem mecanismos para buscar, minerar e recuperar informações relevantes da enorme quantidade de dados disponíveis na web transformando-os em conhecimento (MONTOMOYO; MARTÍNEZ-BARCO; BALAHUR, 2012).

Apesar de já existirem muitos métodos altamente precisos para analisar e extrair conhecimento relevante baseado em dados estruturados (por exemplo, tabelas ou banco de dados), a tarefa de extrair informações úteis com base em dados não estruturados (texto, discurso, etc.), como é o caso das mídias sociais,

ainda, é um importante desafio (LIU; ZHANG, 2012). A solução para essa questão vem sendo procurada por muitos pesquisadores em um subcampo do Processamento de Linguagem Natural³ denominado análise de sentimento (FORTUNY et al., 2012).

4 ANÁLISE DE SENTIMENTO

A análise de sentimento, também conhecida como mineração de opinião, análise de atitude, análise de subjetividade e análise de favorabilidade (PANG; LEE, 2008; LANE; CLARKE; HENDER, 2012; DI CARO; GRELLA, 2013; KONTOPOULOS et al., 2013; MOUTHAMI; DEVI; BHASKARAN, 2013), é uma técnica que busca identificar, extrair e classificar informações subjetivas, como opiniões e sentimentos, sobre diversos temas baseados em textos (FORTUNY et al., 2012; YOON; ELHADAD; BAKKEN, 2013). Para Mostafa (2013), a análise de sentimento pode ser considerada como uma técnica automatizada de descoberta de conhecimento que visa encontrar padrões escondidos em um grande número de dados textuais, assim como os comentários submetidos às mídias sociais. Trata-se de uma tecnologia emergente que identifica a opinião das pessoas em relação a um tema ou objeto, sendo muito útil no monitoramento de mídias sociais para determinar automaticamente o sentimento geral de seus usuários sobre várias questões (HE; ZHA; LI, 2013; MOUTHAMI; DEVI; BHASKARAN, 2013).

O objetivo da análise de sentimento é identificar a opinião das pessoas sobre temas específicos, não se tratando somente de uma estruturação de dados (DI CARO; GRELLA, 2013). Nessa linha de pensamento, Santos et al. (2011)

³ Processamento de Linguagem Natural é uma disciplina de Inteligência Artificial que lida com o tratamento automático da linguagem natural em texto ou discurso, sendo muito empregada, por exemplo, em motores de busca como Google e Bing (MONTORO; MARTÍNEZ-BARCO; BALAHUR, 2012).

esclarecem que a análise de sentimento procura classificar textos não por tópicos e, sim, pelo sentimento ou opinião neles contidos. Seu propósito é o de criar uma base de conhecimento contendo opiniões de uma forma mais organizada e explícita (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

Embora a opinião seja um conceito muito amplo, a análise de sentimento foca, principalmente, sentimentos positivos e negativos (MOREO et al., 2012). Ela é usada para extrair opiniões, sentimentos e subjetividade em textos não estruturados, ou seja, para identificar se as expressões indicam um parecer positivo (favorável) ou negativo (desfavorável) para o assunto (PANG; LEE, 2008). Dessa forma, um trabalho de análise de sentimento pode ser interpretado como uma tarefa de classificação em que cada categoria (positiva ou negativa) representa um sentimento (PRABOWO; THELWALL, 2009). Assim, a análise de sentimento, normalmente, lida com a identificação da polaridade, ou seja, determina se um texto é objetivo ou subjetivo, e se um texto subjetivo contém sentimentos positivos ou negativos, em vez de emoções distintas, como alegria e tristeza, por exemplo, (BAE; LEE, 2012). Kontopoulos et al. (2013) e Mostafa (2013) consideram, também, a opinião neutra dentro da polaridade, ao concordarem que a análise de sentimento é um processo que tem como objetivo determinar se a polaridade de um corpus textual (documento, frase, parágrafo, etc.) tende a ser positivo, negativo ou neutro.

Quanto à forma de classificar os textos, boa parte das abordagens em análise de sentimento é baseada em recursos linguísticos ou na aprendizagem de máquina (KONTOPOULOS et al., 2013; YU; DUAN; CAO, 2013). A abordagem baseada em recursos linguísticos está centrada em listas predeterminadas de palavras positivas e negativas, e a identificação automatizada da polaridade do texto dependerá da frequência dos diferentes tipos de palavras que aparecerem no documento (YU; DUAN; CAO, 2013). A aprendizagem de máquina trata-se da preparação de um classificador de sentimento com base em um conjunto de

treino, ou seja, uma amostra de dados utilizada, para treinar o classificador a distinguir os sentimentos positivos, negativos e neutros, nos textos de forma automatizada (KONTOPOULOS et al., 2013). Na literatura, a abordagem de aprendizagem de máquina tem apresentado melhores resultados quanto à precisão da classificação automatizada dos dados (LIU; ZHANG, 2012; KONTOPOULOS et al., 2013; YANG; YU, 2013). Uma ferramenta baseada nessa abordagem será demonstrada adiante, mas ainda é necessário conhecer as vantagens e limitações dessa técnica.

4.1 Vantagens e limitações da técnica

Atualmente, a análise de sentimentos está recebendo muita atenção por causa da grande variedade de suas aplicações diretas como análises de produtos, serviços, perfil do público, tendências políticas, entre outros (DI CARO; GRELLA, 2013). Até mesmo questões de interesse para a segurança pública ou a qualidade de vida (por exemplo, tráfego urbano, qualidade do ar) podem ser descobertos, monitorados e mitigados, por meio da análise dos fluxos de informação das mídias sociais, detectando padrões e tendências significativas (KAVANAUGH et al., 2012).

O fato de que tantas pessoas se expressam nas mídias sociais sobre qualquer assunto tornam as opiniões menos tendenciosas e, conseqüentemente, mais confiáveis (MONTOTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012). Cáceres (2011) corrobora com essa posição por acreditar que os habitantes do ciberespaço, os “novos cidadãos”, têm muito a dizer, muito a ensinar e propor muitas mudanças. Em decorrência dessas razões, as opiniões expressas na web são cada vez mais consideradas no processo decisório e na obtenção de um *feedback* imparcial sobre diversos assuntos (MONTOTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012). Além disso, trabalhar com informações extraídas das mídias

sociais nem sempre necessita de grandes investimentos por parte da organização (OLIVEIRA; WELCH, 2013).

Algumas limitações da análise de sentimento, baseada em dados extraídos das mídias sociais, também, devem ser observadas. O primeiro problema importante que, ainda, precisa ser resolvido é a falta de recursos substanciais para outros idiomas além do inglês (GLASS; COLBAUGH, 2012; MONTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012; YANG; YU, 2013; ZHANG; HE, 2013). Tais recursos devem ser adaptados para os tipos de texto a serem analisados e para cada uma das línguas envolvidas (MONTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012). Também, há outras questões linguísticas a serem consideradas como a presença de ruídos na comunicação e a presença de linguagem informal como gírias e abreviaturas não oficiais (MOREO et al., 2012; YANG; YU, 2013).

5 TRABALHOS RELACIONADOS

No campo da administração pública, existem poucas pesquisas sobre análise de sentimento ou mineração de opinião. Mas alguns estudos merecem destaque. Fortuny et al. (2012) publicaram um estudo intitulado “*Media coverage in times of political crisis: A text mining approach*” que aborda a cobertura das mídias tradicionais sobre questões políticas na Bélgica em um período de crise no final do ano de 2011. O estudo mediu o sentimento social dos cidadãos em relação aos partidos políticos do país extraíndo dados de notícias de grandes jornais em versões on-line por um período de dez meses. Com a pesquisa, os autores descobriram um viés na cobertura da imprensa tradicional em textos políticos.

Sobkowicz, Kaschesky e Bouchard (2012), em seu artigo “*Opinion mining in social media: Modeling, simulating, and forecasting political opinions*

in the web” investigaram como os conteúdos on-line podem ser explorados para informar aos tomadores de decisão sobre as opiniões dos cidadãos, as tendências emergentes e os impactos potenciais das iniciativas políticas. Os autores concluíram que a mineração de opinião apoia os formuladores de políticas em todas as etapas do ciclo de políticas públicas, antecipando a repercussão das ações perante a opinião pública.

Considerando esses trabalhos, percebe-se que a análise de sentimento, baseada em informações coletadas nas mídias sociais, pode ser uma valiosa fonte de conhecimento sobre a sociedade (TSYTSARAU; PALPANAS, 2012). Em posse desse conhecimento, os agentes públicos podem, segundo Prabowo e Thelwall (2009), analisar a opinião do público no que diz respeito a serviços públicos e questões políticas. Veremos na próxima seção se isso é possível com base em um exemplo prático.

6 A ANÁLISE DE SENTIMENTO NA PRÁTICA

Para demonstrar como a análise de sentimento pode ser utilizada na prática como um termômetro da opinião pública sobre as ações dos governos, tomamos como exemplo as manifestações populares ocorridas no Brasil em junho de 2013 e identificamos a opinião dos cidadãos brasileiros sobre os fatores que motivaram tais manifestações em dois momentos distintos: seis meses e um ano após o ocorrido. De acordo com a pesquisa Ibope (2013b), os protestos no Brasil foram motivados pela insatisfação dos cidadãos com várias questões de interesse público. Outra pesquisa realizada pelo Ibope (2013a) revela, também, que as mídias sociais foram os canais mais utilizados pelos cidadãos para divulgar e organizar os protestos, bem como para comentar sobre o assunto. O propósito desse exemplo é verificar se a análise de sentimento, realizada após o

ocorrido, pode revelar se os cidadãos continuam insatisfeitos com aquelas questões ou se algo mudou.

6.1 Procedimentos Metodológicos

Este é um estudo aplicado, interdisciplinar e exploratório, que pode ser classificado como qualitativo no que diz respeito à coleta, classificação e análise de dados subjetivos, e quantitativo no que se refere a identificar a frequência de opiniões positivas, negativas e neutras no conjunto de dados (COLLIS; HUSSEY, 2005; RAMBOCAS; GAMA, 2013).

A análise de sentimento exige que o pesquisador siga algumas etapas que envolvem basicamente coleta, codificação e processamento de dados (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012). Na literatura não existe um padrão definido, uma vez que os estudos publicados sobre análise de sentimento apresentam métodos e técnicas bastante diversificadas (LIU, 2010; LI; LIU, 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012; ROBALDO; DI CARO, 2013; YOON; ELHADAD; BAKKEN, 2013; ZHANG; HE, 2013). As etapas adotadas neste estudo, para a aplicação da técnica de análise de sentimento, foram baseadas nos estudos de Tsytarau e Palpanas (2012) e Yoon, Elhadad e Bakken (2013), e consistiram basicamente em:

1. Escolher os temas para análise.
2. Definir os termos de busca.
3. Escolher a fonte de dados.
4. Selecionar aplicativo de mineração de opinião.
5. Coletar dados.
6. Preparar dados.
7. Delimitar o conjunto de treino.
8. Realizar a análise automatizada dos dados; e
9. Validar os resultados.

Os temas (etapa 1) foram escolhidos com base nos principais fatores que motivaram as manifestações ocorridas no Brasil identificados pela pesquisa Ibope (2013b) e pela Plataforma Causa Brasil⁴. De acordo com essas pesquisas, os fatores mais citados como causa das manifestações e escolhidos como temas para este estudo são: transporte público, combate à corrupção, saúde, educação, gastos com a Copa do Mundo FIFA 2014 e Governo Federal.

Os termos de busca (etapa 2) foram selecionados por meio de um conjunto de palavras-chave utilizadas no monitoramento de mídias sociais pela Plataforma Causa Brasil durante as manifestações. O critério utilizado foi o de escolher três termos ligados a cada tema. Esses termos deveriam ser diretamente relacionados aos temas e ao mesmo tempo viáveis para aplicação da técnica. Os termos selecionados estão distribuídos no Quadro 1.

Quadro 1 Termos de pesquisa utilizados na mineração de dados.

Temas	Termos de busca
Transporte público	Transporte público, passe livre, preço das passagens.
Combate à corrupção	Combate à corrupção, CPI, julgamento mensalão.
Educação	Educação, salário professor, escola pública.
Saúde	Saúde, hospitais públicos, posto de saúde.
Gastos com a Copa do Mundo	Gastos com a Copa, obras da Copa, gastos com estádios.
Governo Federal	Governo Federal, Dilma Rousseff, Presidenta Dilma.

Fonte: Elaborado pelo autor.

As fontes de dados escolhidas (etapa 3) foram as mídias sociais mais utilizadas pelos cidadãos brasileiros durante as manifestações em junho de 2013. De acordo com a pesquisa Ibope (2013b), em uma das maiores manifestações

⁴ CAUSA BRASIL, 2014

ocorridas no país, 62% dos manifestantes souberam do evento pelo Facebook, e 75% destes convocaram outras pessoas para participar por meio do Facebook e do Twitter. Assim, essas duas mídias representam as fontes de dados para este trabalho.

A seleção do aplicativo (etapa 4), para realizar a mineração de opinião, baseou-se em alguns requisitos. O aplicativo deveria permitir o tratamento de dados em língua portuguesa e ser capaz de extrair e classificar automaticamente dados do Facebook e do Twitter, a partir dos termos de pesquisa definidos anteriormente, além de ser gratuito ou possuir licença para fins acadêmicos. Foram analisadas as ferramentas DiscoverText⁵, Rapidminer⁶ e Scup⁷. Considerando os requisitos, o aplicativo adotado foi o DiscoverText. Sua escolha foi definida pela viabilidade técnica, disponibilidade gratuita para fins acadêmicos e seu desempenho satisfatório em outros estudos (BEYER, 2012). O DiscoverText é um aplicativo de análise de textos, baseado em nuvem capaz de capturar, filtrar e classificar um grande volume de dados estruturados e não estruturados. Por meio desse aplicativo, é possível personalizar e reutilizar classificadores, baseados em aprendizagem de máquina que permitem o uso combinado de algoritmos e codificação humana, o que aumenta a precisão nos resultados (KONTOPOULOS et al., 2013).

A coleta de dados (etapa 5) foi realizada, por meio do aplicativo DiscoverText, entre os dias 1º e 30 de dezembro de 2013 e 1º a 30 de junho de 2014, utilizando todos os termos de busca definidos no Quadro 1. Após a coleta, foi realizada a preparação dos dados (etapa 6), excluindo-se os ruídos (dados duplicados, *links*, entre outros) e aplicados alguns filtros para garantir que a base de dados fosse composta apenas das menções feitas por cidadãos brasileiros.

⁵ DISCOVERTEXT, 2014

⁶ RAPIDMINER, 2014

⁷ SCUP, 2014

O conjunto de treino (etapa 7) foi delimitado com base em uma amostra aleatória simples (COCHRAN, 2007). Para calcular a dimensão ideal do conjunto de treino, o primeiro passo foi extrair uma amostra de 100 menções sobre cada tema, compondo uma base de dados para teste contendo um total de 600 menções. A cada 60 menções classificadas manualmente, equivalente a 10% do total, foi realizada a classificação automatizada de toda a base de teste e, assim, contabilizados os acertos, até conseguir uma precisão satisfatória. A precisão foi calculada, dividindo-se o número de classificações corretas pelo total de menções (EIRINAKI; PISAL; SINGH, 2012). Após classificadas manualmente 30% das menções, a precisão da classificação automatizada chegou aos 81%. Na literatura, uma ferramenta que gera em média 80% de acertos já é considerada satisfatória (MOSTAFA, 2013; YOON; ELHADAD; BAKKEN, 2013; YU; DUAN; CAO, 2013). Dessa forma, o conjunto de treino considerado para este estudo é de 30% das menções coletadas para cada tema, conforme distribuídos na Tabela 1.

Tabela 1 Cálculo do conjunto de treino.

Tema	Dados coletados em Dezembro de 2013			Dados coletados em Junho de 2014		
	Dados brutos	Dados processados	Conjunto de treino (30%)	Dados brutos	Dados processados	Conjunto de treino (30%)
Transporte público	7.874	795	239	15.194	2.482	745
Combate à corrupção	8.595	1.892	568	10.567	2.138	642
Educação	9.996	1.419	426	13.823	1.984	596
Saúde	5.178	814	245	9.874	1.895	567
Gastos com a Copa do Mundo	6.606	1.737	522	10.423	2.196	659
Governo Federal	13.608	2.642	793	19.746	3.244	974
TOTAL	51.857	9.299	2.793	79.627	13.939	4.183

Fonte: Elaborado pelos autores

Conforme os dados constantes na Tabela 1, percebe-se uma grande redução do número de dados brutos para os dados processados. Isso se deve à aplicação de filtros, durante a fase de preparação dos dados, que eliminou os ruídos dos dados brutos, excluindo, também, dados duplicados e as menções feitas por agências de notícias (mídias tradicionais). Dessa forma, o conjunto de dados sobre cada tema foi composto apenas por menções de cidadãos brasileiros relacionadas aos termos de busca previamente definidos. Considerando os dados processados, foi calculado o conjunto de treino para cada tema na proporção de 30% e realizada a classificação manual desses dados, que foram coletados, aleatoriamente, dentro de cada conjunto de dados. Nessa etapa é necessária a classificação manual de cada conjunto de treino para que se estabeleçam regras e padrões (aprendizagem de máquina) para a classificação automatizada dos dados.

Após o treinamento do aplicativo, foi realizada a classificação automatizada (etapa 8) de todos os conjuntos de dados e gerado o sentimento (positivo, neutro e negativo) sobre cada tema.

Além de conhecer o sentimento dos cidadãos perante cada tema, é muito importante validar esses resultados (etapa 9), calculando o quanto são precisos (COLLIS; HUSSEY, 2005). A precisão foi calculada com base em uma nova amostra aleatória, considerando apenas os dados que foram classificados de forma automatizada. O tamanho da amostra foi de 30% do total de dados analisados, mantendo a mesma proporção do teste realizado no conjunto de treino. Dessa forma, o cálculo da precisão da análise automatizada, também baseado no modelo de Eirinaki, Pital e Singh (2012), deveria ser o mais próximo possível do valor encontrado naquele teste, ou seja, 81%.

Tabela 2 Precisão da análise de sentimento.

Tema	Precisão	
	Dados coletados em 12/2013	Dados coletados em 06/2014
Transporte público	76%	73%
Combate à corrupção	74%	78%
Educação	80%	82%
Saúde	82%	83%
Gastos com a Copa do Mundo	77%	76%
Governo Federal	69%	71%
MÉDIA	76,33%	77,16

Fonte: Elaborado pelos autores.

A média aritmética da precisão de todos os temas foi de 76,33% para os dados coletados em dezembro de 2013 e de 77,16% para os dados coletados em junho de 2014. Dessa forma houve uma perda de 4,66% e 3,84% de precisão nos respectivos dados comparando-os com o resultado obtido no primeiro conjunto de teste (81%). Isso aconteceu porque a precisão da mineração de opinião pode variar de acordo com o assunto pesquisado e a qualidade dos dados (EIRINAKI; PISAL; SINGH, 2012).

7 RESULTADOS E DISCUSSÃO

Após a classificação automatizada dos dados relacionados aos temas que motivaram a realização das manifestações populares de junho de 2013 no Brasil, pode-se observar que, mesmo após um ano, os fatores que motivaram estas manifestações continuam gerando insatisfação nos cidadãos brasileiros, conforme demonstrado no Gráfico 1.

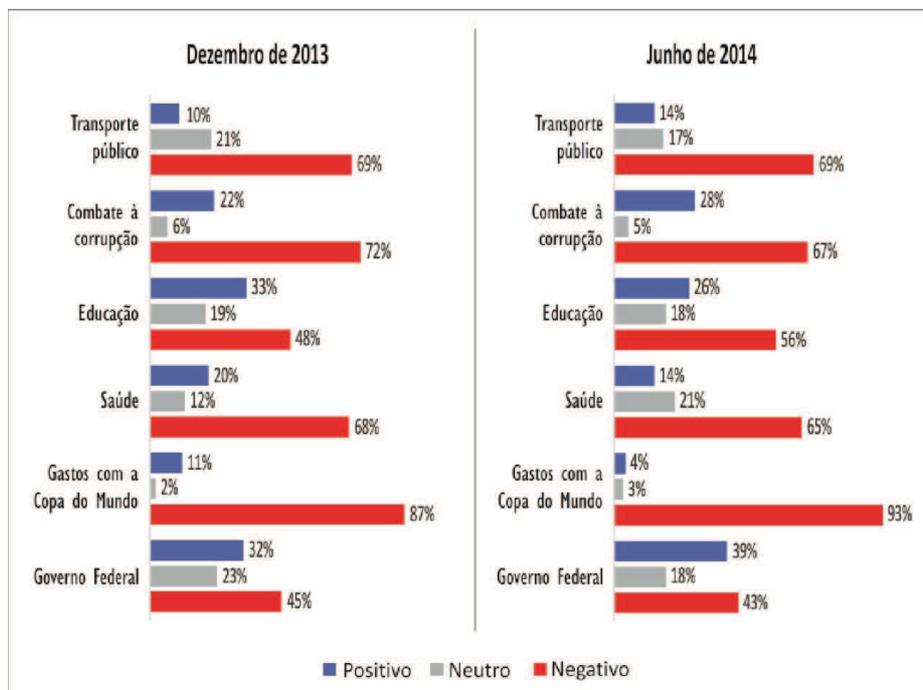


Gráfico 1 Análise de sentimento dos brasileiros baseada nos dados coletados em dezembro de 2013 e junho de 2014.

Fonte: Dados da pesquisa.

Ponderando a análise de sentimento realizada com dados extraídos do Twitter e do Facebook, houve predominância de opiniões negativas em todos os temas tratados. Dentre eles, os **gastos com a Copa do Mundo** foi o que gerou maior número de menções negativas (87% em dezembro de 2013 e 93% em junho de 2014). Por meio da leitura de algumas dessas menções, é possível entender melhor essa tendência negativa. Os cidadãos manifestaram-se nas mídias sociais com uma posição fortemente contra o nível de gastos públicos com os preparativos para a Copa do Mundo FIFA 2014, como podemos perceber na seguinte postagem: “10 bilhões gastos na construção de estádios de futebol. E o povo sem educação, saúde e segurança. E sem futuro” (Usuário A, Twitter, 19

de dezembro de 2013). Por motivos éticos será mantido o anonimato de todos os autores das menções reproduzidas neste trabalho.

O segundo tema com maior número de menções negativas em dezembro de 2013 foi o **combate à corrupção** (72%), caindo para 67% de menções negativas em junho de 2014. Pela leitura dessas menções percebe-se a falta de esperança dos brasileiros no combate efetivo à corrupção, por exemplo:

Os corruptos cassados? Nunca serão! Cidadãos bem informados? Nunca serão! Hospitais bem equipados? Nunca serão! Nunca serão!! Nunca serão!! (Usuário B, Twitter, 22 de dezembro de 2013).

Na sequência de menções negativas, no primeiro gráfico seguem praticamente empatados os temas **transporte público** (69%) e **saúde** (68%). Seis meses depois foi mantido o número de menções negativas sobre o tema transporte público, enquanto as menções negativas sobre o tema saúde tiveram ligeira queda, chegando a 65%. Esses resultados foram impulsionados por menções como:

Estou muito revoltada com a saúde Pública. Como pode na cidade de Tupã não ter uma UTI Neonatal???? Minha filha teve que ser levada às pressas para outra cidade.... O que acontece prefeito??? Ah sim, acabei de me lembrar; toda verba que recebem seria somente para fazer praças, shows, carnaval.... (Usuário C, Facebook, 16 de dezembro de 2013).

Este transporte público do DF está um caos, hoje devido a superlotação do ônibus, teve uma discussão entre passageiros por falta de espaço (Usuário D, Twitter, 15 de dezembro de 2013).

Os temas que apresentaram maior número de menções positivas em dezembro de 2013 foram **educação** (33%) e **Governo Federal** (32%). Isso se manteve em junho de 2014 e as menções positivas sobre o tema educação tiveram uma queda de 7% enquanto no tema Governo Federal houve uma alta de 7%. Esse, também, foi o tema com o maior número de menções neutras em

dezembro de 2013. Em junho de 2014 o tema com maior número de menções neutras foi a saúde. Estas menções foram marcadas por comentários que não expressaram claramente uma posição positiva ou negativa sobre o tema, como: “Presidente Dilma Rousseff acaba de chegar em Porto Alegre” (Usuário E, Twitter, 19 de dezembro de 2013).

Como foi demonstrado na Tabela 2, o tema que apresentou menor precisão, após classificados todos os dados, foi o **Governo Federal**. Por meio da análise criteriosa de algumas menções que foram classificadas erroneamente, percebe-se que a precisão da ferramenta foi afetada por falsos-positivos, falsos-negativos e menções com alto nível de ironia, por exemplo: “Odeio gente ignorante que critica essas bolsas do governo federal. É ser muito mesquinho e ignorante” (Usuário F, Twitter, 14 de junho de 2014).

Entre os seis temas, o que apresentou maior precisão na mineração de opinião automatizada foi a **saúde**. Analisando as menções coletadas sobre esse tema, percebe-se um número razoável de frases contendo opiniões claras e objetivas, por exemplo: “Os hospitais públicos do Brasil são horríveis, entra governo e sai governo, continua a mesma coisa...” (Usuário G, Twitter, 14 de dezembro de 2013)”. Assim, conjuntos de dados que apresentam boa coerência textual podem aumentar a precisão da classificação automatizada. Por outro lado, um conjunto de dados compostos por expressões que contenham ironia, gírias e erros de ortografia podem reduzir consideravelmente a confiabilidade dos resultados (MONTORO; MARTÍNEZ-BARCO; BALAHUR, 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

8 TRABALHOS FUTUROS

Existem, basicamente, dois principais campos para futuras pesquisas sobre análise de sentimento na administração pública. O primeiro é sobre o

desenvolvimento de ferramentas mais precisas que, também, sejam flexíveis ao processar dados em outros idiomas além do inglês (GLASS; COLBAUGH, 2012; MONTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012; YANG; YU, 2013; YOON; ELHADAD; BAKKEN, 2013; ZHANG; HE, 2013). O segundo campo consiste na descoberta de novas formas de aplicar a análise de sentimento em casos reais buscando novas soluções para problemas estratégicos da administração pública. Um caminho promissor é avaliar o quanto a análise de sentimento pode contribuir para promover uma gestão pública mais participativa, estabelecendo uma relação dialógica entre Estado e sociedade.

Trata-se de aplicar a análise de sentimento de forma que possibilite aos governos considerarem a opinião dos cidadãos no processo decisório, contribuindo para uma gestão pública mais democrática.

9 CONCLUSÃO

Este estudo apresentou as mídias sociais como um canal ativo em que os usuários manifestam suas opiniões sobre os mais variados temas, inclusive, sua satisfação ou insatisfação com as ações do governo. Diferentemente dos canais fechados como, por exemplo, telefone e *e-mail*, as informações postadas nas mídias sociais podem ser disseminadas muito rapidamente, contribuindo para a formação da opinião pública, principalmente, sobre temas de grande repercussão. O desafio, no entanto, era aplicar um método de processamento e análise de dados que pudesse gerar conhecimento por meio de dados não estruturados. A proposta apresentada como forma de superar esse desafio foi a análise de sentimento.

A revisão teórica sobre mídias sociais e análise de sentimento, juntamente com o exemplo de aplicação, demonstraram que é possível a utilização das informações postadas pelos cidadãos em mídias sociais como uma

ferramenta de mensuração da opinião pública sobre vários temas. Para exemplificar a aplicação da análise de sentimento nos campos da política e da administração pública, foi apresentado um caso prático tomando como exemplo a opinião dos cidadãos sobre os principais temas que motivaram as manifestações populares ocorridas no Brasil em junho de 2013 em dois períodos posteriores. Os resultados demonstraram que os cidadãos brasileiros, um ano após as manifestações, continuam emitindo opiniões negativas associadas aos fatores que os motivaram a protestar, principalmente os relacionados aos gastos públicos com a Copa do Mundo FIFA 2014, o combate à corrupção e o transporte público.

Diante disso, a análise de sentimento pode contribuir para revelar a opinião dos cidadãos quanto à atuação do governo e aos serviços públicos, identificando aqueles que possuem maior demanda de ampliação e/ou melhorias. A partir daí os governos podem definir prioridades, traçar estratégias para intervir nos problemas, e agir visando atender ao interesse público. Estas ações podem gerar maior nível de satisfação entre os cidadãos, o que, conseqüentemente, contribui na prevenção de conflitos. Principalmente quando os “holofotes” estão voltados para o país em razão da Copa do Mundo FIFA 2014, as eleições presidenciais 2014 e a preparação para os Jogos Olímpicos 2016, ou seja, um palco perfeito para protestos.

Embora os resultados deste estudo tenham apresentado uma precisão razoável na análise automatizada dos dados, algumas limitações devem ser observadas. Primeiramente, estes resultados revelam a opinião dos cidadãos que são usuários das mídias sociais investigadas e que fizeram comentários sobre os temas definidos neste estudo durante dois períodos distintos de 30 dias. Dessa forma, é inadequado generalizar os números aqui apresentados em situações que ultrapassem esse espaço temporal e universo de usuários.

REFERÊNCIAS

- BACHNER, J.; HILL, K. W. Advances in public opinion and policy attitudes research. **Policy Studies Journal**, Urbana, v. 42, p. S51-S70, May 2014.
- BAE, Y.; LEE, H. Sentiment analysis of twitter audiences: measuring the positive or negative influence of popular twitterers. **Journal of the American Society for Information Science and Technology**, New York, v. 63, n. 12, p. 2521-2535, Nov. 2012.
- BEYER, Y. Using discovertext for large scale twitter harvesting. **Microform & Digitization Review**, London, v. 41, n. 3-4, p. 121-125, Dec. 2012.
- BONSÓN, E. et al. Local e-government 2.0: social media and corporate transparency in municipalities. **Government Information Quarterly**, Amsterdam, v. 29, n. 2, p. 123-132, Apr. 2012.
- CÁCERES, J. G. Las nuevas tecnologías de información y comunicación y las políticas culturales en México : ingeniería en comunicación social del servicio de redes sociales facebook. **Intercom**, São Paulo, v. 34, p. 175-196, Nov. 2011.
- CAMBRIA, E. et al. New avenues in opinion mining and sentiment analysis. **IEEE Intelligent Systems**, Los Alamitos, v. 28, n. 2, p. 15-21, Apr. 2013.
- CAUSA BRASIL. Disponível em: < <http://www.causabrasil.com.br>>. Acesso em: 14 mar. 2014.
- CHEN, T. M. How networks changed the world. **IEEE Network**, New York, v. 25, n. 6, p. 2-3, Nov. 2011.
- CHUN, S. A.; LUNA-REYES, L. F. Social media in government. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 441-445, Oct. 2012.
- COCHRAN, W. G. **Sampling techniques**. New York: John Wiley & Sons, 2007.
- COLLIS, J.; HUSSEY, R. **Pesquisa em administração: um guia prático para alunos de graduação e pós-graduação**. 2. ed. Porto Alegre: Bookman, 2005.

DI CARO, L.; GRELLA, M. Sentiment analysis via dependency parsing. **Computer Standards & Interfaces**, Amsterdam, v. 35, n. 5, p. 442-453, Sept. 2013.

DISCOVERTEXT. Disponível em: <<http://www.discovertext.com>>. Acesso em: 22 mar. 2014.

EIRINAKI, M.; PISAL, S.; SINGH, J. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, New York, v. 78, n. 4, p. 1175-1184, July 2012.

FERRO, E. et al. Policy making 2.0: from theory to practice. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 359-368, Oct. 2013.

FORTUNY, E. J. D. et al. Media coverage in times of political crisis: a text mining approach. **Expert Systems with Applications**, New York, v. 39, n. 14, p. 11616-11622, Oct. 2012.

GLASS, K.; COLBAUGH, R. Estimating the sentiment of social media content for security informatics applications. **Security Informatics**, New York, v. 1, n. 1, p. 3, 2012.

GUIMARÃES, J.; AMORIM, A. P. **A corrupção da opinião pública: uma defesa republicana da liberdade de expressão**. São Paulo: Boitempo, 2013.

HABERMAS, J. **Direito e democracia: entre facticidade e validade**. Rio de Janeiro: Tempo Brasileiro 2003.

HE, W.; ZHA, S.; LI, L. Social media competitive analysis and text mining: a case study in the pizza industry. **International Journal of Information Management**, Guildford, v. 33, n. 3, p. 464-472, June 2013.

HOFMANN, S. et al. What makes local governments' online communications successful? Insights from a multi-method analysis of facebook. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 387-396, Oct. 2013.

IBOPE. 89% dos manifestantes não se sentem representados por partidos. **Instituto Brasileiro de Opinião Pública e Estatística**, Rio de Janeiro, abr. 2013. Disponível em: <<http://www.ibope.com.br/pt-br/noticias/Paginas/89-dos-manifestantes-nao-se-sentem-representados-por-partidos.aspx>>. Acesso em: 11 out. 2013a.

_____. Veja pesquisa completa do Ibope sobre os manifestantes. **G1**, São Paulo, 24 jun. 2013. Disponível em: <<http://g1.globo.com/brasil/noticia/2013/06/veja-integra-da-pesquisa-do-ibope-sobre-os-manifestantes.html>> . Acesso em: 12 nov. 2013b.

KAVANAUGH, A. L. et al. Social media use by government: from the routine to the critical. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 480-491, Oct. 2012.

KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4065-4074, Aug. 2013.

LANE, P. C. R.; CLARKE, D.; HENDER, P. On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data. **Decision Support Systems**, Amsterdam, v. 53, n. 4, p. 712-718, Nov. 2012.

LEWIŃSKI, M.; MOHAMMED, D. Deliberate design or unintended consequences: the argumentative uses of Facebook during the Arab spring. **Journal of Public Deliberation**, Ohio, v. 8, n. 1p. 01-13, Apr. 2012.

LI, G.; LIU, F. Application of a clustering method on sentiment analysis. **Journal of Information Science**, Cambridge, v. 38, n. 2, p. 127-139, Apr. 2012.

LIM, M. Clicks, cabs, and coffee houses: social media and oppositional movements in Egypt, 2004-2011. **Journal of Communication**, New York, v. 62, n. 2, p. 231-248, Feb. 2012.

LIU, B. **Sentiment analysis and subjectivity**. 2. ed. Chicago: Handbook of Natural Language, 2010.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining text data**. Chicago: Springer, 2012. Cap. 13, p. 415-463.

MERGEL, I. A framework for interpreting social media interactions in the public sector. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 327-334, Oct. 2013.

MERGEL, I. The social media innovation challenge in the public sector. **Information Polity: the international journal of government & democracy in the information age**, Washington, v. 17, n. 3-4, p. 281-292, Dec. 2012.

MONTOYO, A.; MARTÍNEZ-BARCO, P.; BALAHUR, A. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. **Decision Support Systems**, Amsterdam, v. 53, n. 4, p. 675-679, Nov. 2012.

MOREO, A. et al. Lexicon-based comments-oriented news sentiment analyzer system. **Expert Systems with Applications**, New York, v. 39, n. 10, p. 9166-9180, Aug. 2012.

MOSSBERGER, K.; WU, Y.; CRAWFORD, J. Connecting citizens and local governments? Social media and interactivity in major U.S. cities. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 351-358, Oct. 2013.

MOSTAFA, M. M. More than words: Social networks' text mining for consumer brand sentiments. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4241-4251, Aug. 2013.

MOUTHAMI, K.; DEVI, K. N.; BHASKARAN, V. M. Sentiment analysis and classification based on textual reviews. In: INTERNATIONAL CONFERENCE ON, 2013, [S.l.]. **Proceedings...** [S.l.]: [Information Communication and Embedded Systems, 2013. p. 271-276.

NUNOMURA, E. A imprensa, o twitter e as eleições de 2010 no Brasil. **Intercom**, São Paulo, v. 36, n. 1, p. 103-126, June 2013.

O'CALLAGHAN, D. et al. **Online social media in the syria conflict:** encompassing the extremes and the in-Betweens. Cornell: Cornell University, 2014.

OLIVEIRA, G. H. M.; WELCH, E. W. Social media use in local government: linkage of technology, task, and organizational context. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 397-405, Aug. 2013.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, Washington, v. 2, n. 1-2, p. 1-135, Abr. 2008.

PAPACHARISSI, Z.; OLIVEIRA, M. D. F. Affective news and networked publics: the rhythms of news storytelling on #Egypt. **Journal of Communication**, New York, v. 62, n. 2, p. 266-282, Sept. 2012.

PENNEY, J. Motivations for participating in ‘viral politics’: a qualitative case study of twitter users and the 2012 US presidential election. **Convergence: the international journal of research into new media technologies**, Toronto, 2014. Disponível em: < <http://con.sagepub.com/content/early/2014/05/06/1354856514532074>>. Acesso em: 14 jun. 2014.

PERRIN, A. J.; MCFARLAND, K. Social theory and public opinion. **Annual Review of Sociology**, Palo Alto, v. 37, n. 1, p. 87-107, 2011.

PRABOWO, R.; THELWALL, M. Sentiment analysis: a combined approach. **Journal of Informetrics**, Oxford, v. 3, n. 2, p. 143-157, 2009.

RAMBOCAS, M.; GAMA, J. **Marketing research: the role of sentiment analysis**. Universidade do Porto, Faculdade de Economia do Porto. 2013

RAPIDMINER. Disponível em: < <https://rapidminer.com>>. Acesso em: 22 fev. 2014.

ROBALDO, L.; DI CARO, L. Opinion mining-ML. **Computer Standards & Interfaces**, Amsterdam, v. 35, n. 5, p. 454-469, 2013.

SANTOS, L. M. et al. Twitter, análise de sentimento e desenvolvimento de produtos: quanto os usuários estão expressando suas opiniões? **Prisma.com**, Belo Horizonte, n. 13, p. 01-12, 2011.

SNEAD, J. T. Social media use in the U.S. executive branch. **Government Information Quarterly**, Amsterdam, v.30, n. 1, p. 56-63, 2013.

SCUP. Disponível em: < <http://www.scup.com/>>. Acesso em: 28 fev. 2014.

SOBACI, M. Z.; KARKIN, N. The use of twitter by mayors in Turkey: tweets for better public services? **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 417-425, Oct. 2013.

SOBKOWICZ, P.; KASCHEKY, M.; BOUCHARD, G. Opinion mining in social media: modeling, simulating, and forecasting political opinions in the web. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 470-479, Oct. 2012.

STEENKAMP, M.; HYDE-CLARKE, N. The use of facebook for political commentary in South Africa. **Telematics and Informatics**, Amsterdam, v. 31, n. 1, p. 91-97, Feb. 2014.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, Oxford, v. 24, n. 3, p. 478-514, May 2012.

YANG, D.-H.; YU, G. A method of feature selection and sentiment similarity for Chinese micro-blogs. **Journal of Information Science**, Cambridge, v. 39, n. 4, p. 429-441, Aug. 2013.

YOON, S.; ELHADAD, N.; BAKKEN, S. A practical approach for content mining of tweets. **American Journal of Preventive Medicine**, New York, v. 45, n. 1, p. 122-129, July 2013.

YU, Y.; DUAN, W.; CAO, Q. The impact of social and conventional media on firm equity value: a sentiment analysis approach. **Decision Support Systems**, Amsterdam, v. 55, n. 4, p. 919-926, Nov. 2013.

ZHANG, C. et al. Sentiment analysis of Chinese documents: from sentence to document level. **Journal of the American Society for Information Science and Technology**, New York, v. 60, n. 12, p. 2474-2487, Sept. 2009.

ZHANG, P.; HE, Z. A weakly supervised approach to Chinese sentiment classification using partitioned self-training. **Journal of Information Science**, Cambridge, v. 39, n. 6, p. 815-831, Dec. 2013.

ARTIGO 2

A análise de sentimento em mídias sociais como instrumento para as práticas da Gestão Social¹

RESUMO

Este estudo procura descrever como a análise de sentimento pode contribuir para as práticas da Gestão Social colaborando com a consolidação de um modelo de gestão pública mais democrático. Este é um estudo aplicado, interdisciplinar e exploratório que pode ser classificado como qualitativo e quantitativo. O estudo consistiu na revisão das principais formulações teóricas acerca do tema, na pesquisa empírica e no confronto entre os conceitos e aplicações, envolvendo a Gestão Social, a análise de sentimento e as estratégias de utilização de mídias sociais adotadas pelos governos. A aplicação da análise de sentimento foi demonstrada por meio de dados extraídos do Twitter sobre alguns dos principais programas sociais implementados no Brasil pelo Governo Federal nos últimos anos. Ao analisar os resultados observados nesse exemplo, juntamente com os conceitos presentes na literatura, foi possível identificar que a análise de sentimento pode contribuir para as práticas da Gestão Social no contexto da estratégia de rede.

Palavras-chave: Análise de sentimento. Mineração de opinião. Mídias sociais. Gestão Social.

¹ Um resumo deste estudo, elaborado em coautoria com Paulo Henrique de Souza Bermejo, José Roberto Pereira e Pâmela Aparecida dos Santos, foi aprovado condicionalmente (Anexo I), em sua versão na língua inglesa, para ser apresentado na *14th European Conference on Research Methodology for Business and Management Studies* entre os dias 11 e 12 de junho de 2015.

1 INTRODUÇÃO

O surgimento das mídias sociais alterou drasticamente a maneira com que os indivíduos se comunicam (KONTOPOULOS et al., 2013). Os cidadãos estão cada vez mais utilizando essas mídias para se comunicarem com familiares, amigos, empresas e governos (KAVANAUGH et al., 2012). Sabendo disso, os governos têm adotado, gradativamente, as mídias sociais como um mecanismo que permite estreitar relações com o público (BONSÓN et al., 2012; GUSTAFSSON, 2012; ABDELSALAM et al., 2013; HOFMANN et al., 2013; MERGEL, 2013b; SNEAD, 2013). Mas o desafio é mensurar os resultados dessa comunicação, bem como a opinião dos cidadãos perante a atuação do governo (MERGEL, 2013a).

A adoção de técnicas inovadoras para a análise de tais mídias está criando expectativas sobre as futuras oportunidades de inovação e o surgimento de ferramentas que possibilitem utilizar essas mídias como fonte de conhecimento (CRIADO; SANDOVAL-ALMAZAN; GIL-GARCIA, 2013). Uma técnica promissora é a análise de sentimento, também conhecida como mineração de opinião, que tem o propósito de identificar opiniões polarizadas (positivas, negativas e neutras) sobre determinados temas em bases de dados não estruturadas (SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

Nessa perspectiva, o objetivo deste estudo é investigar como a técnica de análise de sentimento pode contribuir para as práticas da Gestão Social, que é entendida aqui como uma ação gerencial dialógica que busca converter fluxos comunicacionais em ações e decisões políticas. Trata-se de identificar o sentimento social em relação a temas que envolvem o interesse público, tanto estatal quanto não estatal e, dessa forma, estabelecer uma possibilidade, mesmo que preliminar, de participação da sociedade civil no processo decisório da *res publica*. Propõe-se, assim, responder a seguinte questão: Como a técnica de

análise de sentimento, baseada na opinião dos cidadãos extraída de mídias sociais, pode contribuir para as práticas da Gestão Social? Para responder essa questão, foram utilizadas ferramentas de mineração de opinião capazes de identificar a opinião dos cidadãos expressa em mídias sociais, sobre alguns dos principais programas sociais mantidos no Brasil pelo Governo Federal na atualidade.

Pretende-se, basicamente, preencher as lacunas evidenciadas nos estudos de Sobaci e Karkin (2013) e Snead (2013) no que se refere a identificar se as mídias sociais possuem potencial para aumentar a participação social e fortalecer a democracia e, também, as lacunas identificadas por Oliveira, Bermejo e Santos (2015) no que diz respeito a avaliar se a análise de sentimento pode contribuir para promover uma gestão pública participativa, estabelecendo uma relação dialógica entre Estado e sociedade.

Para tanto, este artigo está estruturado da seguinte forma: na próxima seção, são apresentadas as estratégias de utilização de mídias sociais na administração pública. Na terceira seção, é delimitado o potencial participativo oferecido pelas tecnologias da informação e comunicação. Na quarta seção, são explicados os principais conceitos sobre Gestão Social no âmbito das relações entre Estado e sociedade. Na quinta seção, apresenta-se a técnica de análise de sentimento. A metodologia que foi aplicada no estudo empírico está detalhada na sexta seção. Na sétima seção, são apresentados os resultados e finalmente o estudo é concluído na oitava seção.

2 ESTRATÉGIAS DE UTILIZAÇÃO DE MÍDIAS SOCIAIS NA ADMINISTRAÇÃO PÚBLICA

A adoção estratégica de ferramentas de mídia social pelos governos, como *blogs*, *microblogs*, *wikis* e redes sociais eletrônicas, está mudando o

panorama de órgãos públicos e burocracias em todo o mundo (MERGEL, 2013a). Nos EUA, as mídias sociais são utilizadas pelo governo com o objetivo de promover a transparência, a participação e a colaboração, que são características predominantes da *Open Government Initiative* do governo Obama (BERTOT; JAEGER; HANSEN, 2012; CHUN; LUNA-REYES, 2012; MERGEL, 2012; CRIADO; SANDOVAL-ALMAZAN; GIL-GARCIA, 2013; MEIJER; THAENS, 2013; MERGEL, 2013a; MERGEL, 2013b; MOSSBERGER; WU; CRAWFORD, 2013; SNEAD, 2013). Outros estudos revelam que as mídias sociais estão sendo utilizadas pelos governos de vários países como China (ZHENG, 2013), Turquia (SOBACI; KARKIN, 2013), México (PICAZO-VELA; GUTIÉRREZ-MARTÍNEZ; LUNA-REYES, 2012), Suécia (GUSTAFSSON, 2012), Holanda (BEKKERS; EDWARDS; KOOL, 2013) e, também, outros países da União Europeia (BONSÓN et al., 2012). O foco da utilização das mídias sociais pelos governos varia bastante, desde *accountability* até segurança pública e prevenção de tragédias naturais (KAVANAUGH et al., 2012; MEIJER; THAENS, 2013).

Nos últimos anos, as pesquisas sobre o processo de adoção das mídias sociais pelos governos demonstram a aplicação de diferentes estratégias. Mergel (2012) e Meijer e Thaens (2013) apresentam quatro tipos de estratégias: estratégia *push*, em que as mídias sociais são utilizadas como canal de transmissão (cidadãos como receptores de informações); estratégia *pull*, que considera as mídias sociais um canal de entrada do cidadão (cidadãos como sensores do governo); estratégia de *rede*, em que as mídias sociais atuam como canal interativo (cidadãos como coprodutores de políticas governamentais, também conhecida como “coprodução do bem público”); e a estratégia de *transação*, onde há a prestação direta de serviços governamentais por meio de aplicativos de mídia social (cidadãos como parceiros nas atividades do governo).

Para Chun e Luna-Reyes (2012), a utilização das mídias sociais pelos governos pode ser considerada uma inovação tecnológica, bem como um agente

transformador gerando, portanto, o engajamento cívico por meio de iniciativas de participação social que promovem decisões mais democráticas, temas que serão abordados na próxima seção.

3 PARTICIPAÇÃO SOCIAL E A PROMOÇÃO DA DEMOCRACIA

Cançado, Pereira e Tenório (2013) acreditam que a democracia representativa no Brasil poderá abrir mais espaço à participação política do cidadão à medida que os processos de ação comunicativa se ampliem na sociedade. Nesse sentido, faz-se necessário buscar formas de ultrapassar o nível existente de democracia (DAHL, 2001), por meio da democracia participativa, que pode ser definida como “um sistema piramidal, com democracia direta na base e democracia por delegação nos outros níveis” (PAULA, 2005, p. 160). Assim, a ampliação da democracia representativa rumo à democracia participativa está ao alcance da sociedade, principalmente com o surgimento das novas tecnologias de comunicação (CANÇADO; PEREIRA; TENÓRIO, 2013).

Atualmente, as tecnologias da informação e comunicação, em especial as ferramentas baseadas na internet, têm o potencial de contribuir, significativamente, na promoção da participação social em grande escala, rompendo as barreiras da democracia representativa (CHUN; LUNA-REYES, 2012; CRIADO; SANDOVAL-ALMAZAN; GIL-GARCIA, 2013; SOBACI; KARKIN, 2013). Essas tecnologias contribuem para a associação dos cidadãos em prol de objetivos comuns, que segundo Habermas (2003), eleva a sociedade a um nível em que é possível que a mesma seja ouvida pelos governantes.

A defesa da participação social não implica uma deliberação pública constante sobre todos os temas, impondo aos cidadãos um envolvimento permanente e exclusivo na esfera pública, pois “o importante é que as pessoas tratem das questões centrais e que as informações fundamentais estejam

disponíveis” (GUIMARÃES; AMORIM, 2013, p. 130). O desafio é construir um modelo em que o todo esteja representado nas partes, “superando a fragmentação geoespacial e funcional, típica das burocracias clássicas” (KEINERT, 2007, p. 97). Este é um desafio que já havia sido lançado por Alexis de Tocqueville, em sua obra *A democracia na América*, publicada originalmente em 1835, quando afirmou que “até agora, não se descobriu forma política alguma que favoreça igualmente o desenvolvimento e a prosperidade de todas as classes que compõem a sociedade” (TOCQUEVILLE, 1998, p. 180). Na atualidade, surge um campo do conhecimento que propõe uma nova forma de gestão para preencher essa lacuna secular: a Gestão Social.

4 GESTÃO SOCIAL

De acordo com Cançado, Pereira e Tenório (2013), a Gestão Social tem como protagonista a sociedade nas relações com o Estado e o mercado. Trata-se de um processo gerencial dialógico em que a decisão é compartilhada entre as partes envolvidas, ou seja, “o adjetivo social qualificando o substantivo gestão é percebido como o espaço privilegiado de relações sociais onde todos têm direito à fala, sem nenhum tipo de coação” (TENÓRIO, 2006, p. 1146).

Para compreender esse universo da Gestão Social é necessário, a princípio, conhecer melhor alguns termos-chave que são frequentemente relacionados ao tema na literatura, como: sociedade civil, esfera pública, fluxos comunicacionais, cidadania deliberativa e saberes alternativos (HABERMAS, 2003; TENÓRIO, 2006; KEINERT, 2007; TENÓRIO, 2008; NATIVIDADE; PEREIRA; OLIVEIRA, 2011; CANÇADO; PEREIRA; TENÓRIO, 2013; GUIMARÃES; AMORIM, 2013).

De acordo com Habermas (2003, p. 100), a *sociedade civil* é composta por “movimentos, organizações e associações, os quais captam os ecos dos

problemas sociais que ressoam nas esferas privadas, condensam-nos e os transmitem, a seguir, para a esfera pública política”. Para o autor, a sociedade civil pode ter opiniões próprias que são capazes, em alguns casos, de influenciar decisões políticas. Nessa mesma linha, Cançado, Pereira e Tenório (2013, p. 102) afirmam que “é na sociedade civil que se institucionalizam os discursos de interesse público capazes de solucionar problemas, pois exercem influência sobre a formação institucionalizada da opinião e da vontade”.

Quanto à *esfera pública*, Habermas (2003) considera que esta se trata de uma rede de comunicação de conteúdos, tomadas de posição e opiniões onde os fluxos comunicacionais baseados na linguagem natural se condensam em opiniões públicas específicas. Keinert (2007) chama a atenção para uma questão importante ao considerar que, apesar da esfera pública ser historicamente vinculada ao Estado, suas raízes estão na própria sociedade. Dessa forma, cabe à esfera pública reforçar a pressão exercida pelos problemas e não somente identificá-los, buscando problematizá-los e dramatizá-los até que sejam assumidos pelos governantes (HABERMAS, 2003). Sob essa perspectiva, a esfera pública seria um espaço de intermediação entre Estado, sociedade e mercado que deve identificar, compreender e propor soluções para os problemas da sociedade (TENÓRIO, 2008; CANÇADO; PEREIRA; TENÓRIO, 2013).

Os *fluxos comunicacionais* da esfera pública estão muito ligados aos domínios da vida privada e fazem com que a sociedade civil consiga identificar novos problemas antes mesmo que os agentes estatais (HABERMAS, 2003). Desse modo, a Gestão Social, baseada em uma ação comunicativa, seria uma forma de ampliar as possibilidades de emancipação social (CANÇADO; PEREIRA; TENÓRIO, 2013). Nos termos de HABERMAS (2003, p. 22), essa rede comunicacional forma “arenas nas quais pode acontecer uma formação mais ou menos racional da opinião e da vontade acerca de matérias relevantes para toda a sociedade”. Portanto, a formação de um sistema público de comunicação é

fundamental para renovar a base discursiva da soberania popular e dar voz para aqueles que são social ou culturalmente marginalizados e desfavorecidos (GUIMARÃES; AMORIM, 2013). Nesse contexto, Habermas (2003, p. 84) afirma que “a integração de uma sociedade altamente complexa não se efetua através de um sistema paternalista que ignora o poder comunicativo do público de cidadãos”.

A *cidadania deliberativa* é um processo de reflexão e decisão participativo, baseado no entendimento entre as partes, para tomada de decisão dentro da esfera pública (TENÓRIO, 2008; CANÇADO; PEREIRA; TENÓRIO, 2013). Em linhas gerais, a cidadania deliberativa significa que a legitimidade das decisões deve ter origem em “processos de discussão orientados pelos princípios da inclusão, do pluralismo, da igualdade participativa, da autonomia e do bem comum” (TENÓRIO, 2008, p. 160).

Os *saberes alternativos* são complementares ao conhecimento especializado, pois o fato do público ser composto por leigos e “a comunicação pública se dar numa linguagem compreensível a todos não significa necessariamente um obscurecimento das questões essenciais ou das razões que levam a uma decisão” (HABERMAS, 2003, p. 107).

Somente as terminologias apresentadas não são capazes de esgotar o amplo universo em que a Gestão Social está inserida. Mas são suficientes para delimitar um conceito que viabilize a proposta deste estudo:

A Gestão Social se situa no espaço de interseção das relações entre Estado, Mercado e Sociedade e torna-se capaz de converter os *fluxos comunicacionais* das *esferas públicas* em ações e decisões políticas a partir de sua organização na *sociedade civil* [*cidadania deliberativa*]. Sendo assim, a Gestão Social é uma ação gerencial dialógica voltada para o interesse público não estatal e para a realização do bem comum, tendo como tarefas na sociedade civil sistematizar os *saberes alternativos*, estruturar avaliações técnicas especializadas voltadas para o bem comum, fortalecer a esfera pública municiando-a de informações de qualidade,

dentre outras (CANÇADO; PEREIRA; TENÓRIO, 2013, p. 103, grifo nosso).

Após esclarecidos os princípios que envolvem a Gestão Social, resta saber como aplicar esses conceitos. Cançado, Pereira e Tenório (2013) destacam a necessidade de se criar instrumentos de gestão que possam efetivar as propostas da Gestão Social. Para Guimarães e Amorim (2013, p. 138), ao se propor “a formação de uma opinião pública democrática como horizonte, ela convida a imaginação democrática a criar as condições de possibilidade de sua máxima realização no contexto histórico em que se exerce”.

Nessa circunstância, este estudo propõe testar a técnica de análise de sentimento como uma possível contribuição para se estabelecer uma relação dialógica entre Estado e sociedade civil sob a ótica da Gestão Social.

5 ANÁLISE DE SENTIMENTO

A análise de sentimento é um subcampo do Processamento de Linguagem Natural que se preocupa em extrair, classificar e analisar opiniões sobre diversos temas em grandes volumes de dados textuais (FORTUNY et al., 2012; YOON; ELHADAD; BAKKEN, 2013). Também conhecida como mineração de opinião, análise de atitude, análise de subjetividade e análise de favorabilidade (PANG; LEE, 2008; SANTOS et al., 2011; LANE; CLARKE; HENDER, 2012; DI CARO; GRELLA, 2013; KONTOPOULOS *et al.*, 2013), a análise de sentimento tem o objetivo de identificar a opinião das pessoas sobre temas específicos a partir de textos, classificando-os não por tópicos e, sim, pelo sentimento ou opinião (SANTOS et al., 2011; DI CARO; GRELLA, 2013). O propósito é criar uma base de conhecimento contendo opiniões (positivas, negativas ou neutras) sobre temas previamente definidos em bases de dados não estruturadas (PANG; LEE, 2008; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012).

Atualmente, a técnica de análise de sentimento é aplicada sob duas abordagens principais: A primeira consiste na classificação automática da base de dados com base no número de ocorrências de termos pertencentes a um léxico contendo palavras previamente classificadas como positivas ou negativas (GHIASSI; SKINNER; ZIMBRA, 2013). A segunda abordagem se baseia na classificação manual de uma amostra de dados (ou conjunto de treino) que servirá de parâmetro para a classificação automática de toda a base de dados com o auxílio de algoritmos de aprendizagem de máquina (PANG; LEE; VAITHYANATHAN, 2002; PANG; LEE, 2008). A vantagem da segunda abordagem em relação à primeira, é que, ao classificar amostras de dados, serão incorporadas ao classificador palavras pertencentes ao mesmo domínio e contexto da base de dados, que será submetida à classificação automatizada, o que proporciona maior precisão nos resultados (LIU; ZHANG, 2012).

A técnica da análise de sentimento pode ser aplicada, baseando-se em diversos tipos de textos, on-line e off-line, mas este estudo concentra-se na análise de textos extraídos especificamente de mídias sociais, pois estas representam um poderoso instrumento de interação entre governos e cidadãos (KHAN et al., 2014). Dessa maneira, a análise de sentimento se apresenta como uma técnica muito útil no monitoramento de mídias sociais porque revela automaticamente o sentimento geral dos usuários dessas mídias em relação a temas específicos, determinando se estes são vistos positiva ou negativamente (MOUTHAMI; DEVI; BHASKARAN, 2013).

6 METODOLOGIA

O presente estudo trata-se de uma pesquisa aplicada, interdisciplinar e exploratória que pode ser classificada como, qualitativa quanto ao processo de classificação dos dados e quantitativa ao identificar a frequência de opiniões

positivas, negativas e neutras contidas na base de dados (COLLIS; HUSSEY, 2005; RAMBOCAS; GAMA, 2013). Os procedimentos metodológicos adotados foram baseados nos estudos de Yoon, Elhadad e Bakken (2013) e Oliveira, Bermejo e Santos (2015) e constituem-se nas seguintes etapas: definir objeto de estudo; escolher a fonte de dados; definir palavras-chave para mineração de dados; selecionar aplicativo de mineração de opinião; coletar e preparar os dados; delimitar o conjunto de treino; realizar a classificação automatizada; e validar os resultados.

Para verificarmos se a técnica de análise de sentimento pode refletir a opinião pública e, assim, trazer contribuições para as práticas da Gestão Social, tomamos como objeto de estudo a opinião dos cidadãos expressas em mídias sociais sobre os principais programas sociais mantidos atualmente no Brasil pelo Governo Federal. Foram selecionados para análise os quatro programas sociais² considerados pela imprensa brasileira como “bandeiras” do atual governo (CARAZZAI; BRITTO, 2014), sendo os programas:

- Bolsa Família: programa de transferência direta de renda que beneficia famílias em situação de pobreza e de extrema pobreza em todo o país.
- Minha Casa, Minha Vida: programa de financiamento habitacional que oferece descontos, subsídios e redução do valor de seguros habitacionais facilitando a aquisição da casa própria.
- Mais Médicos: programa que busca melhoria do atendimento aos usuários da saúde pública por meio de investimentos em infraestrutura dos hospitais e unidades de saúde, além de levar médicos brasileiros e estrangeiros para regiões onde há escassez e ausência de profissionais; e

² Mais informações sobre cada um desses programas sociais estão disponíveis na seção “Planos e Programas” do Portal Brasil, disponível em <<http://www.brasil.gov.br/planos-e-programas>>. Acesso em: 25 mar. 2014.

- Programa Nacional de Acesso ao Ensino Técnico em Emprego (Pronatec): Esse programa tem como objetivo ampliar a oferta de cursos gratuitos de educação profissional e tecnológica no Brasil.

A fonte de dados foi estabelecida baseada em alguns critérios. Inicialmente deveria ser uma mídia social popular no Brasil. As postagens dos usuários deveriam ser, predominantemente, em forma de texto, pois o insumo da análise de sentimento é a comunicação textual. E, por fim, deveria haver viabilidade técnica para a mineração e análise dos dados. Dessa forma, a mídia social escolhida como fonte de dados foi o Twitter. O Twitter é um *microblog* que pode ser acessado por meio de sua *homepage*³ ou por meio de dispositivos móveis onde seus usuários compartilham informações e opiniões por meio de mensagens curtas, ou *tweets*, limitados a 140 caracteres (DELLER, 2011; LARSSON; MOE, 2011; BAE; LEE, 2012). Em virtude dessa limitação, os *tweets* tendem a ser mais concisos, expressando a opinião dos usuários de forma mais objetiva, o que torna o processamento destes dados mais eficaz do que as postagens longas permitidas em outras mídias sociais (KONTOPOULOS et al., 2013).

Após a escolha da fonte de dados, o próximo passo foi definir as palavras-chave para a realização da mineração de dados. Visando conseguir coletar o maior número possível de postagens sobre cada um dos quatro programas sociais escolhidos como objeto deste estudo, as palavras-chave utilizadas na mineração de dados foram definidas em dois padrões distintos baseados nas regras de busca avançada do Twitter. O primeiro padrão partiu da utilização dos próprios nomes dos quatro programas sociais entre aspas, pois isso garante a busca apenas da expressão exata reduzindo ruídos nos dados. O

³ TWITTER, 2014

segundo padrão foi definido por meio das *hashtags*⁴ oficiais utilizadas no compartilhamento de informações referentes a estes programas. Dessa maneira, as palavras-chave utilizadas na mineração de dados foram definidas como no Quadro 1:

Quadro 1 Palavras-chave utilizadas na mineração de dados.

Programas Sociais	Hashtags
“Bolsa Família”	#BolsaFamília
“Minha Casa, Minha Vida”	#MinhaCasa MinhaVida
“Mais Médicos”	#Mais Médicos
“Pronatec”	#Pronatec

Fonte: Elaborado pelos autores.

O aplicativo selecionado para realização deste estudo foi o DiscoverText⁵. Outros aplicativos como Rapidminer⁶ e Scup⁷ foram analisados, mas a escolha do DiscoverText foi definida pela viabilidade técnica, disponibilidade gratuita para fins acadêmicos e seu desempenho em outros estudos (BEYER, 2012; OLIVEIRA; BERMEJO; SANTOS, 2015). O DiscoverText é um aplicativo de análise de textos, baseado em nuvem capaz de capturar, filtrar e classificar um grande volume de dados estruturados e não estruturados, que, além de facilitar a captura de dados de mídia social, fornece ferramentas analíticas bem integradas para explorar e codificar os dados (BEYER, 2012). Por meio desse aplicativo, é possível personalizar e reutilizar classificadores baseados em aprendizagem de máquina que permitem o uso combinado de algoritmos e codificação humana, o que aumenta a precisão nos resultados (KONTOPOULOS et al., 2013).

⁴ *Hashtag*: Palavra ou frase que precedida pelo símbolo # classifica ou categoriza o texto que o acompanha (como em um *tweet*) (MERRIAM-WEBSTER, 2014).

⁵ DISCOVERTEXT, 2014

⁶ RAPIDMINER, 2014

⁷ SCUP, 2014

A coleta de dados foi realizada entre os dias 3 de julho e 3 de agosto de 2014, por meio do aplicativo DiscoverText, utilizando-se as palavras-chave definidas no Quadro 1. Após a coleta, os dados foram preparados, aplicando-se filtros para eliminação de ruídos como *tweets* em outros idiomas, *links*, dados duplicados, entre outros (FORTUNY et al., 2012).

A dimensão do conjunto de treino, considerada para este trabalho, foi definida em 30% do total de *tweets* coletados sobre cada programa social que compõe o objeto de estudo, pois Oliveira, Bermejo e Santos (2015) identificaram que a classificação manual de 30% dos dados coletados é suficiente para se obter uma precisão satisfatória na classificação automatizada com o aplicativo DiscoverText. Assim, o conjunto de treino foi calculado conforme a Tabela 1.

Tabela 1 Cálculo do conjunto de treino

Programas Sociais	Dados brutos	Dados preparados	Conjunto de treino (30%)
Bolsa Família	12.598	2.033	610
Minha Casa, Minha Vida	15.537	2.525	758
Mais Médicos	16.641	2.342	703
Pronatec	14.236	3.276	983
TOTAL	59.012	10.176	3.054

Fonte: Elaborado pelos autores.

De acordo com a Tabela 1, houve uma grande redução entre os dados brutos e os dados preparados. Isso se deve à aplicação de filtros que eliminaram *tweets* duplicados e outros ruídos nos dados brutos além da exclusão de postagens feitas por agências de notícias, resultando em um conjunto de dados composto apenas por *tweets* de cidadãos brasileiros.

A próxima etapa foi realizar a classificação manual do conjunto de treino. Nessa etapa foram classificados manualmente cada conjunto de treino, para estabelecer regras e padrões (aprendizagem de máquina), para que o aplicativo possa classificar automaticamente todos os dados. Terminada a classificação manual de cada conjunto de treino, foi realizada a classificação automatizada de todos os conjuntos de dados, gerando o sentimento (positivo, negativo e neutro) sobre cada um dos programas sociais.

Após a classificação automatizada, foi realizada a validação dos resultados para saber o quanto são precisos (COLLIS; HUSSEY, 2005; OLIVEIRA; BERMEJO; SANTOS, 2015). A precisão dos resultados foi calculada, com base em uma amostra aleatória simples (COCHRAN, 2007) dos dados classificados de forma automatizada. O tamanho da amostra manteve a mesma proporção do conjunto de treino, ou seja, 30% dos dados. O cálculo da precisão foi baseado no modelo de Eirinaki, Pisal e Singh (2012) dividindo-se o número total de classificações corretas pelo número total de cada amostra, como demonstrado na Tabela 2.

Tabela 2 Precisão da análise automatizada dos dados.

Programas Sociais	A	B	C	D
	Dados classificados de forma automatizada	Amostra para cálculo da precisão ($A \times 30\%$)	Dados classificados corretamente	Precisão (%) $C \div B \times 100$
Bolsa Família	1.423	427	318	74,47%
Minha Casa, Minha Vida	1.767	530	441	83,20%
Mais Médicos	1.639	492	385	78,25%
Pronatec	2.293	688	592	86,04%

Fonte: Elaborado pelos autores.

A coluna A da Tabela 2 representa os dados classificados de forma automatizada e é resultado da subtração entre os dados preparados e o conjunto de treino que foi classificado manualmente (vide Tabela 1). A coluna B se refere a amostra para o cálculo da precisão que foi definida na proporção de 30% dos dados da coluna A. A coluna C corresponde ao número de dados da amostra classificados corretamente pelo aplicativo. A coluna D representa a precisão da classificação automatizada que foi obtida pela divisão do número de classificações corretas (coluna C) pelo total de cada amostra (coluna B).

A precisão alcançada pelo aplicativo utilizado neste estudo pode ser considerada satisfatória, pois, na literatura, uma ferramenta que gera em média 80% de acertos já é considerada satisfatória (MOSTAFA, 2013; YOON; ELHADAD; BAKKEN, 2013; YU; DUAN; CAO, 2013). Mesmo que alguns conjuntos de dados tenham apresentado precisão inferior a 80%, pode-se afirmar que a confiabilidade alcançada pela média aritmética dos quatro conjuntos de dados apresentados é de 80,49%.

7 RESULTADOS E DISCUSSÃO

Os resultados da classificação automatizada dos *tweets*, em relação aos programas sociais, revelaram que nos programas “Pronatec” e “Minha Casa, Minha Vida” houve predominância de opiniões positivas, enquanto os *tweets* sobre os programas “Mais Médicos” e “Bolsa Família” foram majoritariamente negativos, como demonstrado no Gráfico 1.

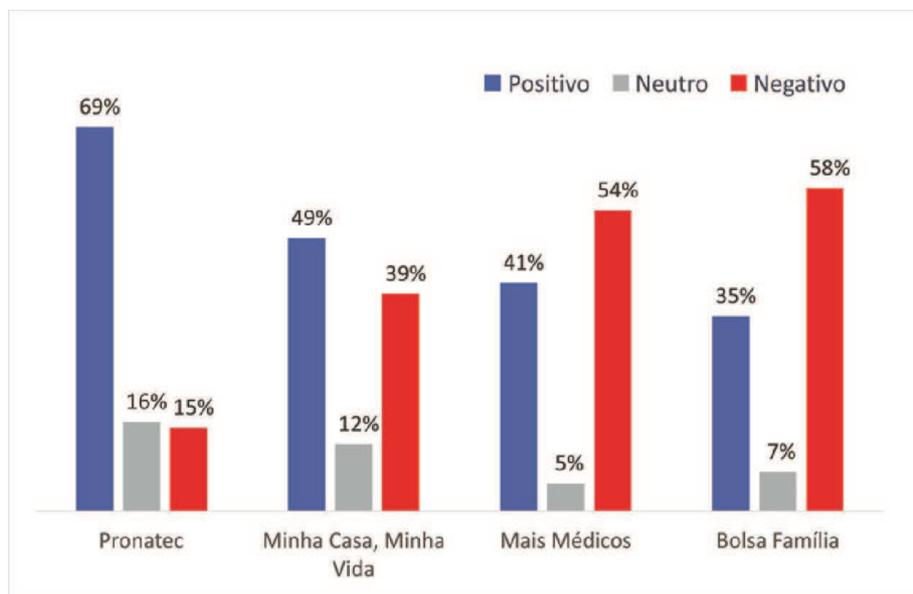


Gráfico 1 Opinião polarizada dos brasileiros em relação aos quatro programas sociais.
Fonte: Dados da pesquisa.

Conforme os resultados da análise de sentimento, o programa social que apresentou o maior número de *tweets* positivos foi o “Pronatec” (69%). Esse foi, também, o programa com maior número de *tweets* neutros (16%). O programa “Minha Casa, Minha Vida” apresentou uma tímida vantagem de *tweets* positivos (49%) em relação às opiniões negativas (39%).

Com predominância de opiniões negativas, os programas “Mais Médicos” e “Bolsa Família” atingiram 54% e 58% de *tweets* negativos, respectivamente. Estes programas sociais, também, foram aqueles que apresentaram menor precisão na classificação automatizada. Um dos motivos é a presença de falsos-positivos ou falsos-negativos classificados pelo aplicativo. Isso se deve a *tweets* com alta carga de ironia, gírias ou erros ortográficos (MONTORO; MARTÍNEZ-BARCO; BALAHUR, 2012). Alguns exemplos de *tweets* classificados neste estudo constam no Quadro 2.

Quadro 2 Exemplo de *tweets* classificados pela polaridade.

Polaridade	Tweets
Positivo	“somos todos #Pronatec! haha amoooo!” (Usuário A, Twitter, 17/07/2014).
Negativo	“o Minha Casa, Minha Vida, do jeito que é feito, não consegue solucionar o déficit habitacional do Brasil” (Usuário B, Twitter, 24/07/2014).
Neutro	“Vale a pena ler pelo depoimento de médicos e pacientes: Como está o programa Mais Médicos no RS um ano depois” (Usuário C, Twitter, 27/07/2014).
Falso-positivo	“Vou dormir pra ter energia e aumentar o PIB amanhã... e ajudar a pagar a bolsa família...” (Usuário D, Twitter, 23/07/2014).

Fonte: Elaborado pelo autor

Apurados os resultados da análise de sentimento sobre os quatro programas sociais em questão, ainda resta saber se essa técnica pode realmente contribuir para as práticas da Gestão Social. Essa questão pode ser respondida, por meio do confronto entre os conceitos e aplicações, envolvendo as estratégias de utilização de mídias sociais pelos governos (MERGEL, 2012; MEIJER; THAENS, 2013), a Gestão Social (TENÓRIO, 2006; CANÇADO; PEREIRA; TENÓRIO, 2013) e a técnica de análise de sentimento (PANG; LEE, 2008; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012), que foram apresentados anteriormente neste estudo e agora distribuídos no Quadro 3.

Quadro 3 Relação entre estratégias de utilização de mídias sociais, Gestão Social e análise de sentimento.

Estratégias de utilização de mídias sociais (MERGEL, 2012; MEIJER; THAENS, 2013)	Atende aos princípios da Gestão Social? (TENÓRIO, 2006; CANÇADO; PEREIRA; TENÓRIO, 2013)	Há possibilidade de aplicação por meio da análise de sentimento? (PANG; LEE, 2008; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012)
Push	Não	Não
Pull	Não	Sim
Rede	Sim	Sim
Transação	Sim	Não

Fonte: Elaborado pelo autor

De acordo com o Quadro 3, a análise de sentimento pode contribuir para as práticas da Gestão Social, promovendo a coprodução de políticas governamentais (estratégia de rede). Como visto anteriormente, a Gestão Social é uma ação gerencial dialógica voltada para o interesse público não estatal capaz de converter os fluxos comunicacionais (saberes alternativos) das esferas públicas em ações e decisões políticas para a realização do bem comum (CANÇADO; PEREIRA; TENÓRIO, 2013). Analisando as quatro estratégias sobre o processo de adoção de mídias sociais pelos governos descritas por Mergel (2012) e Meijer e Thaens (2013), sob a ótica da Gestão Social, percebe-se que as estratégias *push* (cidadãos como receptores de informações) e *pull* (cidadãos como sensores do governo) não atendem aos requisitos para as práticas da Gestão Social, pois são formas unilaterais de comunicação. Já as estratégias de *rede* (cidadãos como coprodutores de políticas governamentais) e de *transação* (cidadãos como parceiros nas atividades do governo) estão diretamente relacionadas aos princípios da Gestão Social, por estabelecerem uma relação dialógica entre Estado e sociedade.

De forma análoga, ao analisar a possibilidade de aplicação destas estratégias, por meio da técnica de análise de sentimento, percebe-se que as estratégias *push* e de *transação* não podem ser aplicadas utilizando-se essa técnica. Na estratégia *push*, a comunicação se dá de forma unilateral, do governo para os cidadãos, por meio da disseminação de informações nos perfis oficiais dos órgãos governamentais, não sendo viável a utilização da análise de sentimento nesse caso, uma vez que as publicações oficiais tendem a ser informativas e não opinativas. No caso da estratégia de *transação*, a análise de sentimento não pode contribuir para a prestação direta de serviços públicos de forma eletrônica. Já as estratégias *pull* e de *rede* são possíveis de serem aplicadas por meio da técnica de análise de sentimento. Na estratégia *pull*, a análise de sentimento pode ser um sensor da opinião dos cidadãos sobre diversos

temas relacionados ao governo. Quanto à estratégia de *rede*, a análise de sentimento pode favorecer a coprodução de políticas governamentais. Mas como isso é possível? Conforme já visto, as mídias sociais são instrumentos eficazes no exercício da liberdade de expressão, que é fundamental para que os cidadãos participem mais ativamente da gestão pública e tenham a capacidade de pressionar e direcionar as decisões do governo (HABERMAS, 2003). Verificou-se, também, que essas mídias possibilitam a partilha de um grande volume de informações e opiniões que podem ser utilizadas no processo decisório governamental (MERGEL, 2013b). O desafio, nesse caso, é transformar uma grande quantidade de dados não estruturados em conhecimento que possa influenciar decisões governamentais (MERGEL, 2013a), caracterizando um modelo de Gestão Social nos termos de Caçado, Pereira e Tenório (2013).

Desse modo, os resultados obtidos na análise de sentimento sobre os programas sociais no Brasil revelam opiniões públicas informais que idealmente se formaram em estruturas de uma esfera pública não desvirtuada pelo poder (HABERMAS, 2003), utilizando-se o Twitter como uma fonte de dados alternativa e independente que possibilitou aos cidadãos exercerem sua liberdade de expressão (DAHL, 2001). Assim, a análise de sentimento pode atuar na interseção das relações entre Estado e sociedade sistematizando e convertendo esses fluxos comunicacionais em conhecimento capaz de incorporar a opinião da sociedade civil no corpo político (GUIMARÃES; AMORIM, 2013).

8 CONCLUSÃO

Este estudo propôs que a análise de sentimento, realizada por meio da mineração de dados extraídos de mídias sociais, seja utilizada como uma ferramenta de mensuração da opinião pública visando contribuir com as práticas da Gestão Social. A revisão teórica revelou que as mídias sociais possuem um

grande potencial para aumentar a participação social e fortalecer a democracia, mas ainda se fazia necessário uma forma de transformar a grande quantidade de dados não estruturados em conhecimento.

Os resultados aferidos pela aplicação da análise de sentimento foram capazes de identificar com precisão satisfatória o número de opiniões positivas, negativas e neutras referentes aos quatro programas sociais adotados como objeto de estudo. Foram revelados dois programas sociais com predominância de opiniões positivas (“Pronatec” e “Minha Casa, Minha Vida”) e outros dois com maioria de opiniões negativas (“Mais Médicos” e “Bolsa Família”). Esses resultados revelaram que a técnica de análise de sentimento pode contribuir com as práticas de Gestão Social mensurando a opinião da sociedade civil sobre diversos temas de interesse público. Uma vez que é na sociedade civil que se institucionalizam os discursos capazes de influenciar as decisões políticas (CANÇADO; PEREIRA; TENÓRIO, 2013), a análise de sentimento pode contribuir para que a opinião dos cidadãos seja considerada nas decisões governamentais. Trata-se de se estabelecer uma gestão pública que não centralize o poder de decisão apenas no Estado e que considere a complexidade das relações entre os atores envolvidos, incentivando diferentes canais de participação. Isso poderá resultar em novos desenhos institucionais, criando-se “organizações administrativas efetivas, permeáveis, a partir da participação popular e com autonomia para operar em favor do interesse público” (PAULA, 2005, p. 159).

No entanto, este estudo apresenta algumas limitações. Primeiramente, os resultados da análise de sentimento se restringem a opinião dos cidadãos durante o período de coleta dos dados. Além disso, esses resultados revelam a opinião apenas dos usuários do Twitter, que não necessariamente são atendidos pelos programas sociais utilizados neste estudo. Por isso, é preciso ponderação ao interpretar esses resultados para que não se estabeleça uma “ditadura da maioria”. Outra limitação é a redução da precisão da classificação automatizada

em textos com alta carga de ironia. A presença de ironia nos textos é, ainda, um dos grandes desafios na análise de sentimento (MONTOYO; MARTÍNEZ-BARCO; BALAHUR, 2012).

Enfim, a análise de sentimento demonstrou ser uma técnica muito útil no contexto da Gestão Social, mas são necessárias futuras pesquisas testando a eficácia da técnica em diferentes situações comparando-a com outras técnicas de pesquisa de opinião, bem como a busca por novas tecnologias que possam dar voz aos cidadãos, contribuindo na transição de uma democracia representativa para a democracia deliberativa. Também, são necessárias pesquisas que se preocupem em buscar novas formas para incorporar as opiniões extraídas das mídias sociais nos processos decisórios governamentais.

REFERÊNCIAS

ABDELSALAM, H. M. et al. Social media in Egyptian government websites: presence, usage, and effectiveness. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 406-416, Oct. 2013.

BAE, Y.; LEE, H. Sentiment analysis of twitter audiences: measuring the positive or negative influence of popular twitterers. **Journal of the American Society for Information Science and Technology**, New York, v. 63, n. 12, p. 2521-2535, Oct. 2012.

BEKKERS, V.; EDWARDS, A.; KOOL, D. de. Social media monitoring: responsive governance in the shadow of surveillance? **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 335-342, Jan. 2013.

BERTOT, J. C.; JAEGER, P. T.; HANSEN, D. The impact of policies on government social media usage: issues, challenges, and recommendations. **Government Information Quarterly**, Amsterdam, v. 29, n. 1, p. 30-40, Jan. 2012.

BEYER, Y. Using discovertext for large scale twitter harvesting. **Microform & Digitization Review**, London, v. 41, n. 3-4, p. 121-125, Dec. 2012.

BONSÓN, E. et al. Local e-government 2.0: social media and corporate transparency in municipalities. **Government Information Quarterly**, Amsterdam, v. 29, n. 2, p. 123-132, Apr. 2012.

CANÇADO, A. C.; PEREIRA, J. R.; TENÓRIO, F. G. **Gestão social: epistemologia de um paradigma**. Curitiba: CRV, 2013.

CARAZZAI, E. H.; BRITTO, P. 'Muda melhor quem sabe mudar', diz Dilma, em clima de campanha. **Folha de São Paulo**, São Paulo, 03 jul. 2014. Disponível em: <<http://www1.folha.uol.com.br/poder/2014/07/1480765-muda-melhor-quem-sabe-mudar-diz-dilma-em-clima-de-campanha.shtml>>. Acesso em: 21 abr. 2014.

CHUN, S. A.; LUNA-REYES, L. F. Social media in government. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 441-445, Oct. 2012.

COCHRAN, W. G. **Sampling techniques**. New York: John Wiley & Sons, 2007.

COLLIS, J.; HUSSEY, R. **Pesquisa em administração: um guia prático para alunos de graduação e pós-graduação**. 2. ed. Porto Alegre: Bookman, 2005.

CRIADO, J. I.; SANDOVAL-ALMAZAN, R.; GIL-GARCIA, J. R. Government innovation through social media. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 319-326, Oct. 2013.

DAHL, R. A. **Sobre a democracia**. Brasília: Editora da UnB, 2001.

DELLER, R. Twittering on: audience research and participation using twitter. **Participations**, Sheffield, v. 8, n. 1, p. 216-245, May 2011.

DI CARO, L.; GRELLA, M. Sentiment analysis via dependency parsing. **Computer Standards & Interfaces**, Amsterdam, v. 35, n. 5, p. 442-453, Setp. 2013.

DISCOVERTEXT. Disponível em: <<http://www.discovertext.com>>. Acesso em: 22 mar. 2014.

EIRINAKI, M.; PISAL, S.; SINGH, J. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, New York, v. 78, n. 4, p. 1175-1184, July 2012.

FORTUNY, E. J. D. et al. Media coverage in times of political crisis: a text mining approach. **Expert Systems with Applications**, New York, v. 39, n. 14, p. 11616-11622, Oct. 2012.

GHIASSI, M.; SKINNER, J.; ZIMBRA, D. Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. **Expert Systems with Applications**, New York, v. 40, n. 16, p. 6266-6282, Nov. 2013.

GUIMARÃES, J.; AMORIM, A. P. **A corrupção da opinião pública: uma defesa republicana da liberdade de expressão**. São Paulo: Boitempo, 2013.

GUSTAFSSON, N. The subtle nature of facebook politics: swedish social network site users and political participation. **New Media & Society**, Milano, v. 14, n. 7, p. 1111-1127, Nov. 2012.

HABERMAS, J. **Direito e democracia: entre facticidade e validade**. Rio de Janeiro: Tempo Brasileiro, 2003.

HOFMANN, S. et al. What makes local governments' online communications successful? Insights from a multi-method analysis of facebook. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 387-396, Oct. 2013.

KAVANAUGH, A. L. et al. Social media use by government: From the routine to the critical. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 480-491, Oct. 2012.

KEINERT, T. M. M. **Administração pública no Brasil: crises e mudanças de paradigmas**. 2. ed. São Paulo: Annablume, 2007.

KHAN, G. F. et al. W. From e-government to social government: Twitter use by Korea's central government. **Online Information Review**, Bradford, v. 38, n. 1, p. 95-113, Jan. 2014.

KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4065-4074, Aug. 2013.

LANE, P. C. R.; CLARKE, D.; HENDER, P. On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data. **Decision Support Systems**, Amsterdam, v. 53, n. 4, p. 712-718, Nov. 2012.

LARSSON, A. O.; MOE, H. Studying political microblogging: twitter users in the 2010 swedish election campaign. **New Media & Society**, Milano, v. 14, n. 5, p. 729-747, Aug. 2011.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining text data**. Chicago: Springer, 2012. Cap. 13, p. 415-463.

MEIJER, A.; THAENS, M. Social media strategies: understanding the differences between North American police departments. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 343-350, 2013.

MERGEL, I. A framework for interpreting social media interactions in the public sector. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 327-334, Dec. 2013a.

MERGEL, I. Social media adoption and resulting tactics in the U.S. federal government. **Government Information Quarterly**, Amsterdam, v. 30, n. 2, p. 123-130, Oct. 2013b.

MERGEL, I. The social media innovation challenge in the public sector. **Information Polity: the international journal of government & democracy in the information age**, Washington, v. 17, n. 3-4, p. 281-292, Oct. 2012.

MERRIAM-WEBSTER. Definition of hashtag. **Merriam-Webster Online Dictionary**, Oxford, 2014. Disponível em: <<http://www.merriam-webster.com/dictionary/hashtag>>. Acesso em: 12 jun. 2014.

MONTOYO, A.; MARTÍNEZ-BARCO, P.; BALAHUR, A. Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. **Decision Support Systems**, Amsterdam, v. 53, n. 4, p. 675-679, Apr. 2012.

MOSSBERGER, K.; WU, Y.; CRAWFORD, J. Connecting citizens and local governments? Social media and interactivity in major U.S. cities. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 351-358, Nov. 2013.

MOSTAFA, M. M. More than words: social networks' text mining for consumer brand sentiments. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4241-4251, Aug. 2013.

MOUTHAMI, K.; DEVI, K. N.; BHASKARAN, V. M. Sentiment analysis and classification based on textual reviews. In: INTERNATIONAL CONFERENCE ON, 2013, [S.l.]. **Proceedings...** [S.l.]: [Information Communication and Embedded Systems, 2013. p. 271-276.

NATIVIDADE, E. A.; PEREIRA, J. R.; OLIVEIRA, V. A. R. Gestão de política pública de trabalho e renda: uma análise do Programa Nacional de Incubadoras de Cooperativas Populares. In: PEREIRA, J. R. (Ed.). **Gestão social de políticas públicas**. Lavras: Editora UFLA, 2011. Cap. 3, p. 93-142.

OLIVEIRA, D. J. S.; BERMEJO, P. H. S.; SANTOS, P. A. Sentiment analysis, social media, and public administration. In: CEMAL, D. et al. (Ed.). **Handbook of research on democratic strategies and citizen-centered E-government services**. Hershey: IGI Global, 2015. Cap. 13, p. 231-249.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, Washington, v. 2, n. 1-2, p. 1-135, Jan. 2008.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment classification using machine learning techniques. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE, 10., 2002, Stroudsburg. **Proceedings...** Stroudsburg: Association for Computational Linguistics, 2002. p.79-86.

PAULA, A. P. P. **Por uma nova gestão pública: limites e potencialidades da experiência contemporânea**. Rio de Janeiro: Editora da FGV, 2005.

PICAZO-VELA, S.; GUTIÉRREZ-MARTÍNEZ, I.; LUNA-REYES, L. F. Understanding risks, benefits, and strategic alternatives of social media applications in the public sector. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 504-511, Oct. 2012.

RAMBOCAS, M.; GAMA, J. **Marketing research: the role of sentiment analysis**. Porto: Universidade do Porto, 2013.

RAPIDMINER. Disponível em: < <https://rapidminer.com>>. Acesso em: 22 fev. 2014.

SANTOS, L. M. et al. Twitter, análise de sentimento e desenvolvimento de produtos: quanto os usuários estão expressando suas opiniões? **Prisma.com**, Belo Horizonte, n. 13, p. 01-12, 2011.

SCUP. Disponível em: < <http://www.scup.com/>>. Acesso em: 28 fev. 2014.

SNEAD, J. T. Social media use in the U.S. executive branch. **Government Information Quarterly**, Amsterdam, v. 30, n. 1, p. 56-63, Jan. 2013.

SOBACI, M. Z.; KARKIN, N. The use of twitter by mayors in Turkey: tweets for better public services? **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 417-425, Oct. 2013.

SOBKOWICZ, P.; KASCHEKY, M.; BOUCHARD, G. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 470-479, Oct. 2012.

TENÓRIO, F. G. A trajetória do Programa de Estudos em Gestão Social (Pegs). **Revista de Administração Pública**, São Paulo, v. 40, n. 6, p. 1145-1162, dez. 2006.

TENÓRIO, F. G. **Tem razão a administração? Ensaios de teoria organizacional**. 3. ed. Ijuí: Editora da Unijuí, 2008.

TOCQUEVILLE, A. D. **A democracia na América**. 4. ed. Belo Horizonte: Itatiaia, 1998.

TWITTER. Disponível em: < <http://www.twitter.com/>>. Acesso em: 12 mar. 2014.

YOON, S.; ELHADAD, N.; BAKKEN, S. A practical approach for content mining of tweets. **American Journal of Preventive Medicine**, New York, v. 45, n. 1, p. 122-129, July 2013.

YU, Y.; DUAN, W.; CAO, Q. The impact of social and conventional media on firm equity value: a sentiment analysis approach. **Decision Support Systems**, Amsterdam, v. 55, n. 4, p. 919-926, Nov. 2013.

ZHENG, L. Social media in Chinese government: drivers, challenges and capabilities. **Government Information Quarterly**, Amsterdam, v. 30, n. 4, p. 369-376, Oct. 2013.

ARTIGO 3

As mídias sociais podem revelar as preferências dos eleitores? Um confronto entre análise de sentimento e pesquisas de opinião tradicionais

RESUMO

Este estudo procura identificar se os resultados obtidos pela aplicação da técnica de análise de sentimento, em dados extraídos de mídias sociais, pode revelar precisamente as preferências políticas dos cidadãos se comparada a pesquisas de opinião tradicionais. Para tanto foram coletados, preparados e analisados 92.441 *tweets*, a partir de termos-chave, relacionados aos candidatos Aécio Neves e Dilma Rousseff, durante o segundo turno das eleições presidenciais do Brasil em 2014, revelando a opinião positiva, negativa e neutra dos cidadãos perante estes candidatos. Posteriormente, os resultados dessa análise foram confrontados com os resultados de seis pesquisas de intenção de votos e de rejeição dos candidatos realizada pelo Instituto de Pesquisas Datafolha. Trata-se de um estudo aplicado, interdisciplinar, exploratório, qualitativo e quantitativo. Os resultados deste estudo demonstraram que a análise de sentimento pode indicar a preferência dos eleitores com aproximação variando entre 1% e 8% em relação às pesquisas tradicionais, apresentando precisão média de 81,05% na análise automatizada dos dados.

Palavras-chave: Análise de sentimento. Mídias sociais. Twitter. Eleições. Pesquisa de opinião.

1 INTRODUÇÃO

As mídias sociais perceptivelmente ampliaram o compartilhamento de informações, opiniões e interesses por parte de seus usuários (KHAN; BASHIR; QAMAR, 2014). Como resultado, uma grande quantidade de conteúdo gerado por estes usuários está disponível e acessível, o que representa uma valiosa fonte de dados para identificar a opinião das pessoas sobre diversos assuntos (CERON et al., 2014; HE; ZHA; LI, 2013).

As opiniões expressas nas mídias sociais estão desempenhando um papel importante que influencia tudo, desde os produtos que as pessoas compram até o candidato presidencial que elas apoiam (EIRINAKI; PISAL; SINGH, 2012). A imprensa, os candidatos e os partidos políticos têm investido em pesquisas de opinião pública para saberem a preferência dos eleitores, mas o crescimento exponencial das mídias sociais levanta a possibilidade de se utilizar, também, a web para explorar as preferências políticas dos cidadãos (TUMASJAN et al., 2011; CERON et al., 2014; KORENEK; ŠIMKO, 2014). Principalmente com o surgimento de novas técnicas e ferramentas para a classificação e análise de grandes volumes de dados, atualmente é possível descobrir sentimentos e opiniões que estão contidos em textos não estruturados (ROBALDO; DI CARO, 2013). Entre os métodos mais populares, destaca-se a análise de sentimento (ou mineração de opinião), que está preocupada com o estudo computacional de opiniões pessoais, avaliações ou sentimentos expressos em textos (LIU, 2010; ZHANG; HE, 2013).

Nos estudos sobre análise de sentimento, *microblogs* como o Twitter têm sido uma das principais fontes de dados em razão de suas características que permitem a troca instantânea de informações, onde os usuários podem expressar seus sentimentos, opiniões e ideias para quem possa estar interessado (BAKLIWAL et al., 2013; YOON; ELHADAD; BAKKEN, 2013; KORENEK; ŠIMKO,

2014; SILVA; HRUSCHKA; HRUSCHKA JÚNIOR, 2014). De acordo com Mohammad et al. (2014), o Twitter tem características únicas que o tornam uma rica fonte de dados para a aplicação da análise de sentimento se comparado a outras mídias sociais.

Nessa perspectiva, o presente estudo foi realizado com o objetivo de aplicar a análise de sentimento, baseada na opinião dos cidadãos expressa no Twitter, como método que pode revelar, precisamente, as suas preferências políticas se comparada a pesquisas de opinião tradicionais. Pretende-se, assim, responder a seguinte questão: A análise de sentimento realizada baseada em dados extraídos do Twitter pode revelar as preferências políticas dos cidadãos como nas pesquisas de opinião pública? Essa questão está diretamente relacionada a uma lacuna identificada na literatura por Vargo et al. (2014) no que se refere a investigar o quanto a análise de sentimento é confiável, por meio do confronto dos resultados da análise automatizada de grandes volumes de dados não estruturados com métodos de pesquisa tradicionais. Para responder essa questão, a técnica de análise de sentimento foi aplicada em dados extraídos do Twitter, relacionados ao segundo turno das eleições presidenciais de 2014 no Brasil, revelando o volume de opiniões positivas, negativas e neutras sobre os candidatos Aécio Neves e Dilma Rousseff. Posteriormente, esses resultados foram confrontados com seis pesquisas eleitorais de intenção de votos e rejeição dos candidatos, realizadas pelo Instituto de Pesquisas Datafolha¹ durante o segundo turno das eleições. Essa comparação demonstrou que a análise de sentimento gerou resultados aproximados aos das referidas pesquisas eleitorais.

Este estudo está estruturado da seguinte forma: na próxima seção, as mídias sociais são apresentadas como uma alternativa às pesquisas de opinião tradicionais para identificar a opinião das pessoas sobre diversos temas. Na

¹ O Instituto de Pesquisas Datafolha é um dos mais importantes institutos de pesquisa de opinião do Brasil. Além de realizar pesquisas eleitorais, o Datafolha faz levantamentos estatísticos sobre esportes, mercado e opinião pública.

terceira seção, são descritas as principais abordagens sobre a técnica de análise de sentimento baseada em dados extraídos de mídias sociais. Na quarta seção, são apresentadas as características do Twitter como fonte de dados para a análise de sentimento. Na quinta seção, são demonstrados alguns casos de aplicação da análise de sentimento em trabalhos relacionados. A metodologia aplicada neste estudo está detalhada na sexta seção. Os resultados são apresentados e discutidos na sétima seção e finalmente o estudo é concluído na oitava seção.

2 PESQUISAS DE OPINIÃO EM MÍDIAS SOCIAIS

As pesquisas de opinião pública tradicionais têm sido discutidas na literatura como uma forma de representar adequadamente os pontos de vista de um público (PERRIN; MCFARLAND, 2011). Há algumas décadas, a única forma de se identificar a opinião pública era por meio das pesquisas de opinião face a face (STEENKAMP; HYDE-CLARKE, 2014). Com a evolução da internet e o surgimento das mídias sociais, pesquisadores de opinião passaram a contar com uma nova fonte inesgotável de informações (BACHNER; HILL, 2014).

As mídias sociais proporcionaram aos usuários da internet um novo espaço para se expressarem compartilhando seus pensamentos e opiniões sobre diferentes temas (KHAN; BASHIR; QAMAR, 2014). Cada vez mais *blogs*, fóruns da web e plataformas de redes sociais, como Twitter, Facebook, YouTube, MySpace e LinkedIn, oferecem ao seu público ambientes interativos que tornam possível o intercâmbio de informações, conhecimentos, opiniões e emoções, possibilitando uma comunicação rápida, aberta e acessível (KORENEK; ŠIMKO, 2014; STEENKAMP; HYDE-CLARKE, 2014). Isso fez com que as mídias sociais se tornassem excelentes fontes de dados para pesquisas de opinião (KONTOPOULOS et al., 2013).

De meados do século XX até os dias atuais, as pesquisas de opinião pública tiveram grandes avanços metodológicos no que diz respeito aos métodos de coleta e análise de dados (PERRIN; MCFARLAND, 2011). Da mesma forma, as técnicas que possibilitam coletar e analisar dados extraídos de mídias sociais, com a finalidade de revelar a opinião das pessoas, também, vêm evoluindo nos últimos anos (YU; DUAN; CAO, 2013). Existem muitas vantagens em utilizar dados de mídias sociais para medir a opinião em relação às pesquisas de opinião tradicionais, e uma delas é que a análise de dados on-line é muito menos dispendiosa do que as pesquisas tradicionais de larga escala (BACHNER; HILL, 2014). E, além disso, a análise de mídias sociais são mais rápidas e permitem o monitoramento contínuo da opinião pública em tempo real (CERON et al., 2014). Já as pesquisas de opinião tradicionais são notoriamente sujeitas a uma série de problemas como baixa taxa de respostas, falta de comprometimento dos respondentes, opções de respostas restritas e temas limitados (PERRIN; MCFARLAND, 2011; MANZA; BROOKS, 2012; CERON et al., 2014). O uso de postagens de mídia social, como um meio de medir a opinião pública, atenua muitos destes problemas, já que as opiniões expressas por seus usuários são espontâneas e, muitas vezes, ricas em conteúdo (BACHNER; HILL, 2014). Mas por mais rigoroso que seja o método aplicado nas pesquisas de opinião tradicionais ou em mídias sociais, pesquisas baseadas em dados subjetivos, como a opinião das pessoas, nunca estarão isentas de incertezas ou mesmo viés (MANZA; BROOKS, 2012). Além disso, há uma limitação sobre a representatividade das mídias sociais a ser considerada, pois os usuários dessas mídias não são necessariamente representativos de toda a população de cidadãos de um país (CERON et al., 2014).

Mesmo assim, as mídias sociais tornaram-se excelentes fontes de dados sobre a opinião das pessoas, pois nelas milhões de pessoas podem expressar suas opiniões sobre qualquer assunto, inclusive, os temas mais comuns tratados pelas

pesquisas de opinião tradicionais, como as eleições (EIRINAKI; PISAL; SINGH, 2012; LIU; ZHANG, 2012). Em decorrência de sua popularidade, as plataformas de mídia social têm sido apontadas atualmente como a principal forma para engajar eleitores, uma vez que elas oferecem novas oportunidades para os cidadãos participarem, discutirem e debaterem no campo político (CHOY et al., 2012; STEENKAMP; HYDE-CLARKE, 2014).

Nesse contexto, o uso crescente dos meios de comunicação social por parte de um público mais variado aumentou muito a possibilidade de se investigar as mídias sociais como dispositivos que permitem explorar as preferências políticas dos cidadãos (CERON et al., 2014). Mas isso demanda a aplicação de tecnologias avançadas que permitam extrair, classificar e analisar opiniões, por meio de dados extraídos de mídias sociais (YOON; ELHADAD; BAKKEN, 2013). Uma ferramenta que tem se destacado na literatura quanto ao cumprimento dessas tarefas é a análise de sentimento (LIU; ZHANG, 2012; SOBKOWICZ; KASCHEKY; BOUCHARD, 2012; ZHANG; HE, 2013).

3 ANÁLISE DE SENTIMENTO EM MÍDIAS SOCIAIS

A análise de sentimento, também conhecida como mineração de opinião é o processo que tem como tarefa básica a classificação automática da polaridade de um corpus textual indicando se o mesmo tende a ser positivo, negativo ou neutro (KONTOPOULOS et al., 2013). A principal função da análise de sentimento é descobrir o que as pessoas pensam e quais são as suas opiniões sobre determinados temas com o auxílio da tecnologia (YANG; YU, 2013). Com tal característica, a análise de sentimento reduz drasticamente a necessidade de se ler grandes quantidades de documentos para extrair opiniões (YU; DUAN; CAO, 2013). Trata-se de uma técnica que procura cumprir o desafio de identificar e extrair informações subjetivas de grandes volumes de dados não

estruturados por meio da combinação de técnicas de mineração de dados, aprendizagem de máquina, processamento de linguagem natural, recuperação da informação e gestão do conhecimento (MOUTHAMI; DEVI; BHASKARAN, 2013; YANG; YU, 2013).

De acordo com Ghiassi, Skinner e Zimbra (2013), as principais abordagens sobre análise de sentimento podem ser classificadas em duas categorias: aprendizagem de máquina e orientação semântica. A técnica de aprendizagem de máquina exige a classificação manual de amostras da base de dados (conjunto de treino) para se construir modelos baseados no texto que servirão de parâmetro na análise automatizada dos dados (aprendizado supervisionado) (GUERRA et al., 2011). A técnica de orientação semântica consiste na aplicação de um conjunto de palavras previamente construído, contendo termos positivos e negativos, de forma que a identificação automatizada da polaridade seja determinada pela frequência desses termos no texto, sem a necessidade de que uma amostra da base de dados seja previamente categorizada (aprendizado não supervisionado) (YU; DUAN; CAO, 2013). Entre as duas abordagens, os classificadores de sentimento, baseados em aprendizagem de máquina, têm apresentado resultados mais precisos e por isso o modelo prevalece na maioria das pesquisas (KONTOPOULOS et al., 2013; YANG; YU, 2013). A desvantagem dos métodos, baseados em aprendizagem de máquina, está voltada principalmente para a classificação manual de uma amostra de cada conjunto de dados para treinar o classificador (KONTOPOULOS et al., 2013). Mas essa classificação é necessária, porque um classificador treinado com documentos opinativos de um domínio, muitas vezes, executa mal a classificação automatizada em documentos de outro domínio (LIU; ZHANG, 2012).

Para Yu, Duan e Cao (2013), há três principais razões para escolher a análise de sentimento como uma abordagem de pesquisa: 1) converte grandes

volumes de dados não estruturados em informações que permitem previsões sobre questões específicas; 2) constrói modelos para agregar a opinião coletiva e revela informações úteis sobre o comportamento da população para previsão de tendências futuras; 3) possibilita recolher informações sobre a reação das pessoas a respeito de produtos, personalidades, campanhas publicitárias, campanhas eleitorais e vários outros temas.

A ascensão das mídias sociais, como *blogs*, *microblogs* e *sites* de redes sociais tem despertado o interesse de vários pesquisadores em explorar a opinião dos usuários dessas mídias por meio da análise de sentimento (MOUTHAMI; DEVI; BHASKARAN, 2013). De acordo com Zhang e He (2013), o fato dos cidadãos estarem cada vez mais utilizando essas mídias para expressarem suas opiniões sobre temas variados (marcas, produtos, candidatos políticos, entre outros) fez com que as mídias sociais se tornassem atualmente uma das principais fontes de dados para a análise de sentimento. O crescimento do número de pesquisas, envolvendo análise de sentimento trouxe progressos significativos para as técnicas de rastreamento de opiniões em conteúdos de mídias sociais, especialmente para aqueles que adotam o Twitter como fonte de dados (YU; DUAN; CAO, 2013).

4 TWITTER COMO FONTE DE DADOS

O Twitter é um microblog em que os usuários postam mensagens em tempo real chamadas de “*tweets*” possuindo extensão máxima de 140 caracteres (KUMAR; SEBASTIAN, 2012; SILVA; HRUSCHKA; HRUSCHKA JÚNIOR, 2014). Desenvolvido em 2006, o Twitter se tornou uma interessante fonte de dados para pesquisa em função de sua popularidade, capacidade de expor em tempo real as opiniões dos cidadãos sobre qualquer assunto e por disponibilizar um grande volume de dados gratuitamente (YANG; YU, 2013; SMAILOVIĆ et al., 2014).

Esse *microblog* se diferencia de outras mídias sociais, principalmente, porque seus dados são públicos, enquanto em outras mídias há a possibilidade de comunicação privada, o que as tornam menos interessantes como fonte de dados para pesquisa (VARGO et al., 2014).

Em decorrência de suas características voltadas para a simplificação da comunicação, o Twitter permite que milhões de usuários postem e compartilhem mensagens curtas, por meio de uma variedade de tecnologias, tais como *smartphones*, *tablets*, computadores e outros dispositivos conectados à internet (ZAPPAVIGNA, 2011; KUMAR; SEBASTIAN, 2012). Outra característica importante do Twitter como fonte de dados é que o limite de caracteres nos *tweets* obriga os usuários a serem mais concisos e, eventualmente, mais expressivos do que em outras mídias sociais, o que torna a análise desses dados mais eficaz (KONTOPOULOS et al., 2013).

Um recurso importante do Twitter, para que os usuários possam ampliar a discussão sobre temas específicos, mesmo com a limitação de caracteres, são as *hashtags*, ou seja, os usuários incluem o caractere “#” em uma palavra ou frase conectando os *tweets* a um tema em particular (LARSSON; MOE, 2011). Em outras palavras, as *hashtags* inscrevem uma palavra-chave em um *tweet* fazendo referência ao tema da mensagem que foi atribuído pelo usuário (ZAPPAVIGNA, 2011). Dessa forma, a utilização das *hashtags* faz com que os *tweets* fiquem visíveis para um público maior e ao mesmo tempo torna mais fácil e rápido a manifestação da opinião de outros usuários sobre o tema em questão (KUMAR; SEBASTIAN, 2012; KORENEK; ŠIMKO, 2014).

Nos últimos anos, o potencial do Twitter como fonte de dados tem chamado a atenção de pesquisadores de vários campos do conhecimento que buscam, nesse *microblog*, obter informações sobre a opinião de seus usuários (TUMASJAN et al., 2011). Visto que a grande quantidade de conteúdo gerado pelos usuários do Twitter tornou-se abundante e de fácil acesso, muitos

pesquisadores passaram a se interessar pelo poder preditivo desse *microblog* (SMAILOVIĆ et al., 2014). Especialmente em previsões do mercado de ações (YU; DUAN; CAO, 2013; SMAILOVIĆ et al., 2014) e na previsão do resultado de eleições (O'CONNOR et al., 2010; TUMASJAN et al., 2011; CHOY et al., 2012; CERON et al., 2014; O'CALLAGHAN et al., 2014), muitos pesquisadores empregam a abordagem da análise de sentimento para identificar e avaliar as opiniões dos usuários expressas em mídias sociais (GHIASSI; SKINNER; ZIMBRA, 2013; SMAILOVIĆ et al., 2014). Se por um lado há um número crescente de estudos que consideram o poder preditivo da análise de sentimento, baseada em dados do Twitter, por outro há, também, na literatura alguns cétricos (GAYO-AVELLO, 2012). As críticas ao modelo têm incentivado pesquisadores a buscarem ferramentas cada vez mais precisas para realizar a análise de sentimento em mídias sociais, alcançando, em alguns casos, resultados com alto nível de precisão (KHAN et al., 2014; KORENEK; ŠIMKO, 2014; MOHAMMAD et al., 2014; SILVA; HRUSCHKA; HRUSCHKA JÚNIOR, 2014).

Para Ghiassi, Skinner e Zimbra (2013), os *tweets* possuem características e propriedades únicas que viabilizam a aplicabilidade e contribuem para a eficácia das mais tradicionais abordagens sobre análise de sentimento. É evidente que a disponibilização de informações em mídias sociais como o Twitter geraram a criação de uma coleção pública inigualável de opiniões sobre qualquer tema global importante (KUMAR; SEBASTIAN, 2012). Assim, o Twitter pode ser considerado uma rica e variada fonte de dados para a aplicação da análise de sentimento (KHAN; BASHIR; QAMAR, 2014).

5 TRABALHOS RELACIONADOS

O estudo de Tumasjan et al. (2011) intitulado “*Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape*” busca identificar

se o conteúdo do Twitter pode ser utilizado para prever resultados de eleições. Foram analisados mais de 100.000 *tweets* que citam partidos políticos antes da eleição federal de 2009 do parlamento nacional na Alemanha. Mais de um terço dos *tweets* indicaram que o Twitter não é utilizado apenas para os usuários emitirem sua opinião política, mas também discuti-las com outros usuários. Os resultados revelaram que o simples número de *tweets* que citam um partido político pode ser considerado uma reflexão plausível da porcentagem de votos, e seu poder preditivo chega perto de pesquisas eleitorais tradicionais. Sobre essa questão, Tumasjan et al. (2011) sugerem que seu método pode ser comparado a pesquisas tradicionais, mas não fazem nenhum confronto de resultados com regras bem fundamentadas.

Tumasjan et al. (2011) foram contestados por Jungherr, Jürgens e Schoen (2012) no estudo “*Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment*”, no qual provaram empiricamente que a frequência de menções dos partidos políticos no Twitter não é um indicador válido do sentimento político ou mesmo dos resultados de eleições, o que inviabiliza a aplicação do método em pesquisas futuras.

Guerra et al. (2011), em seu estudo intitulado “*From Bias to Opinion: A Transfer-Learning Approach to Real-Time Sentiment Analysis*”, analisam o sentimento dos usuários do Twitter sobre dois temas: eleições presidenciais de 2010 no Brasil e a temporada do Campeonato Brasileiro de Futebol em 2010. Nesse estudo os autores não consideram a polaridade dos *tweets*, pois acreditam encontrar um viés (preferência) dos usuários pelos candidatos ou times de futebol a partir de *retweets*. Os resultados do estudo demonstram que observando o viés de apenas 10% dos usuários é possível conseguir um nível de precisão variando de 80% a 90% na previsão de sentimento dos usuários nos

tweets. O método de Guerra et al. (2011) considera cada *retweet* como um endosso do usuário concordando com o *tweet* original. Por outro lado, Xu et al. (2014) acreditam que os *retweets* não revelam se os usuários concordam ou não com os pontos de vista expressos nos *tweets* originais. De acordo com o autor, o número de *retweets* apenas indica a influência de um usuário no processo de difusão de informações.

O estudo “*Sentiment, emotion, purpose, and style in electoral tweets*” de Mohammad et al. (2014) teve como objeto de estudo as eleições presidenciais dos Estados Unidos em 2012. Foram desenvolvidos sistemas de classificadores automáticos supervisionados que utilizam amostras de dados rotulados para detectar o sentimento dos *tweets*. O sistema apresentou precisão máxima de 73,91%. Ao final os autores identificaram que os *tweets* eleitorais transmitem emoções negativas duas vezes mais que emoções positivas.

No presente estudo procurou-se abordar algumas questões não contempladas por estes trabalhos. Em primeiro lugar, foi realizado o confronto entre os resultados obtidos pela aplicação da análise de sentimento com os resultados de pesquisas de opinião tradicionais, o que não foi realizado em nenhum dos trabalhos aqui apresentados. Ao contrário do estudo de Guerra et al. (2011), a análise dos dados proposta nesta pesquisa desconsidera os *retweets*, pois estes não revelam se os usuários concordam ou não com os pontos de vista expressos nos *tweets* originais, como afirmou Xu et al. (2014). E a afirmação de Mohammad et al. (2014) sobre os *tweets* eleitorais transmitirem emoções negativas duas vezes mais que emoções positivas, será colocada à prova.

6 METODOLOGIA

O presente estudo trata-se de uma pesquisa aplicada, interdisciplinar e exploratória que pode ser classificada como qualitativa, no que diz respeito à

classificação dos dados, e quantitativa no que se refere a identificar a frequência de opiniões positivas, negativas e neutras na base de dados (COLLIS; HUSSEY, 2005; RAMBOCAS; GAMA, 2013).

Conforme descrito anteriormente, este estudo tem como propósito comparar os resultados obtidos pela análise de sentimento, realizada a partir de dados extraídos de mídias sociais com os resultados apresentados por uma pesquisa de opinião tradicional. Para tanto, propõe-se confrontar os resultados polarizados (opiniões positivas e negativas) da análise de sentimento com os resultados de seis pesquisas eleitorais de intenção de votos e rejeição dos candidatos, realizada pelo Datafolha (2014), durante o segundo turno das eleições presidenciais no Brasil. Essa comparação é possível em virtude da relação estreita que existe entre opiniões positivas e intenção de votos (aceitação) como também entre opiniões negativas e rejeição (SOMASUNDARAN; WIEBE; RUPPENHOFER, 2008).

Os procedimentos metodológicos, para a realização da análise de sentimento, foram baseados nos estudos de Yoon, Elhadad e Bakken (2013) e Oliveira, Bermejo e Santos (2015), e constituíram-se nas seguintes etapas: definir objeto de estudo; escolher a fonte de dados; definir termos-chave para mineração de dados; selecionar aplicativo de mineração de opinião; coletar e preparar os dados; delimitar o conjunto de treino; realizar a análise de sentimento; e validar os resultados.

O objeto de estudo definido para a realização da análise de sentimento foi o segundo turno das eleições presidenciais do Brasil em 2014. A escolha das eleições no Brasil se deu pelo momento oportuno de vivenciar o processo eleitoral e acompanhar simultaneamente as pesquisas eleitorais e as discussões sobre os candidatos nas mídias sociais. Foi escolhida uma eleição presidencial porque estas envolveram milhões de eleitores, o que, conseqüentemente, pode gerar maior interação nas mídias sociais (LARSSON; MOE, 2011).

A fonte de dados escolhida foi o Twitter, por apresentar uma base de dados mais concisa, fazendo com que os *tweets* possam ser processados de forma mais eficaz do que textos longos presentes em outras mídias sociais (KONTOPOULOS et al., 2013). De acordo com Silva, Hruschka e Hruschka Júnior (2014), o Twitter, também, pode ser utilizado para identificar a aceitação e rejeição dos políticos pela sociedade.

Os termos-chave para mineração de dados foram definidos, de acordo com as regras de busca avançada do Twitter, por meio dos nomes dos candidatos e das principais *hashtags* relacionadas a eles que foram compartilhadas, durante a campanha eleitoral, mantendo a uniformidade no padrão dos termos para os dois candidatos, conforme a Quadro 1.

Quadro 1 Termos-chave utilizados na mineração de dados

Nomes	<i>Hashtags</i>
Dilma Rousseff, dilma, dilmarousseff	#dilma, #dilma13
Aécio Neves, aecio, aecioneves	#aecio, #aecio45

Fonte: Elaborado pelo autor.

A seleção do aplicativo, para realizar análise de sentimento, partiu da análise das ferramentas DiscoverText², Rapidminer³ e Scup⁴. O aplicativo selecionado, para realizar a mineração dos dados, foi o DiscoverText. Trata-se de um aplicativo capaz de coletar, filtrar e classificar um grande volume de dados não-estruturados, além de fornecer ferramentas analíticas integradas para explorar e codificar os dados (BEYER, 2012). A escolha desse aplicativo se deu pela disponibilidade de licença gratuita para fins acadêmicos e seu bom

² DISCOVERTEXT, 2014

³ RAPIDMINER, 2014

⁴ SCUP, 2014

desempenho em outros estudos (BEYER, 2012; OLIVEIRA; BERMEJO; SANTOS, 2015).

A coleta dos dados no Twitter foi realizada, a partir dos termos-chave previamente definidos, utilizando-se o aplicativo DiscoverText entre os dias 07 a 23 de outubro de 2014. A base de dados coletada nesse período foi dividida em seis conjuntos de dados, garantindo-se, por meio de filtros, que os *tweets* constantes em cada um desses conjuntos foram postados nas mesmas datas em que o Datafolha (2014) realizou as entrevistas das pesquisas eleitorais cujos resultados serão utilizados neste estudo para fins de comparação. A distribuição dos conjuntos de dados consta na Tabela 1.

Tabela 1 Distribuição dos conjuntos de dados.

Pesquisa eleitoral Datafolha (2014)			Análise de sentimento		
Pesquisa	Data das entrevistas	Eleitores entrevistados	Conjunto de dados	Coleta de dados	Tweets coletados (dados brutos)
1	07 a 09/10/2014	2.879	1	07 a 09/10/2014	37.817
2	14 e 15/10/2014	9.081	2	14 e 15/10/2014	32.501
3	20/10/2014	4.389	3	20/10/2014	27.839
4	21/10/2014	4.355	4	21/10/2014	34.526
5	22 e 23/10/2014	9.910	5	22 e 23/10/2014	54.037
6	24 e 25/10/2014	19.318	6	24 e 25/10/2014	92.441

Fonte: Elaborado pelo autor.

Após a coleta dos dados, os mesmos foram preparados com a aplicação de filtros para se excluir dados duplicados (*retweets*), dados em outros idiomas, entre outros (FORTUNY et al., 2012; MOHAMMAD et al., 2014; XU et al., 2014).

Em seguida, foi definido conjunto de treino, ou seja, uma amostra dos *tweets* que será submetida à classificação humana para treinar o aplicativo que realizará a classificação automatizada de todo o conjunto de dados (GHIASSI; SKINNER; ZIMBRA, 2013). O conjunto de treino foi calculado baseado em uma amostra aleatória simples (COCHRAN, 2007) de 30% dos dados preparados. Em um estudo similar, Oliveira, Bermejo e Santos (2015) identificaram que um conjunto de treino contendo 30% dos dados é suficiente para se obter uma precisão satisfatória na classificação automatizada utilizando o aplicativo DiscoverText. Assim, o conjunto de treino foi definido como na Tabela 2.

Tabela 2 Conjunto de treino.

Conjunto de dados	Dados Brutos	Dados preparados	Conjunto de treino
1	37.817	6.027	1.810
2	32.501	5.309	1.593
3	27.839	4.390	1.317
4	34.526	4.902	1.471
5	54.037	7.388	2.217
6	92.441	12.196	3.659

Fonte: Elaborado pelo autor.

Uma vez definido o conjunto de treino, foi realizada a classificação manual dessa amostra para treinar o aplicativo. Nessa etapa é necessário que o pesquisador classifique manualmente cada conjunto de treino, para que se estabeleçam regras e padrões (aprendizagem de máquina), para a classificação automatizada dos dados.

Com o aplicativo treinado, todos os conjuntos de dados foram classificados automaticamente, revelando o sentimento positivo, negativo e neutro sobre os candidatos.

Terminada a classificação, os resultados foram validados calculando-se o quanto são precisos (COLLIS; HUSSEY, 2005; OLIVEIRA; BERMEJO; SANTOS, 2015). A precisão foi calculada, baseada em uma nova amostra aleatória de 30% do conjunto de dados, considerando apenas os *tweets*, classificados automaticamente, na qual se dividiu o número de classificações corretas pelo total de cada amostra (EIRINAKI; PISAL; SINGH, 2012; OLIVEIRA; BERMEJO; SANTOS, 2015). Os resultados do cálculo da precisão constam na Tabela 3.

Tabela 3 Precisão da classificação automatizada dos dados.

Conjunto de dados	Dados classificados automaticamente	Dados classificados corretamente	Precisão
1	4.217	3.506	83,13%
2	3.716	3.023	81,34%
3	3.073	2.515	81,85%
4	3.431	2.771	80,75%
5	5.171	4.144	80,14%
6	8.537	6.754	79,11%
MÉDIA			81,05%

Fonte: Elaborado pelo autor.

O cálculo da precisão revelou que a classificação automatizada dos *tweets* teve 81,05% de acertos em média. Esses valores ficaram abaixo da precisão apresentada pela pesquisa Datafolha (2014) que foi de 95%. Os níveis de precisão apresentados demonstram que a análise automatizada dos dados teve um bom desempenho, uma vez que na literatura uma precisão de 80% aplicando-se a técnica de análise de sentimento já é considerada satisfatória (MOSTAFA, 2013; YOON; ELHADAD; BAKKEN, 2013; YU; DUAN; CAO, 2013). Pesquisas recentes apresentaram precisão entre 73,91% a 87,57% na classificação automatizada dos dados (KHAN; BASHIR; QAMAR, 2014; KORENEK; ŠIMKO, 2014; MOHAMMAD et al., 2014; VARGO et al., 2014; OLIVEIRA; BERMEJO; SANTOS, 2015).

7 RESULTADOS E DISCUSSÃO

Assim como nas pesquisas de opinião tradicionais, na maioria das aplicações de análise de sentimento, os resultados são demonstrados por meio de gráficos que possibilitam uma visão concisa, o que facilita a comparação com resultados de outras pesquisas (LIU; ZHANG, 2012). Portanto, os resultados da análise de sentimento – realizada a partir dos seis conjuntos de dados, contendo *tweets* relacionados aos candidatos Aécio Neves e Dilma Rousseff, que foram coletados, durante o segundo turno das eleições presidenciais no Brasil em 2014 – podem ser observados no Gráfico 1.

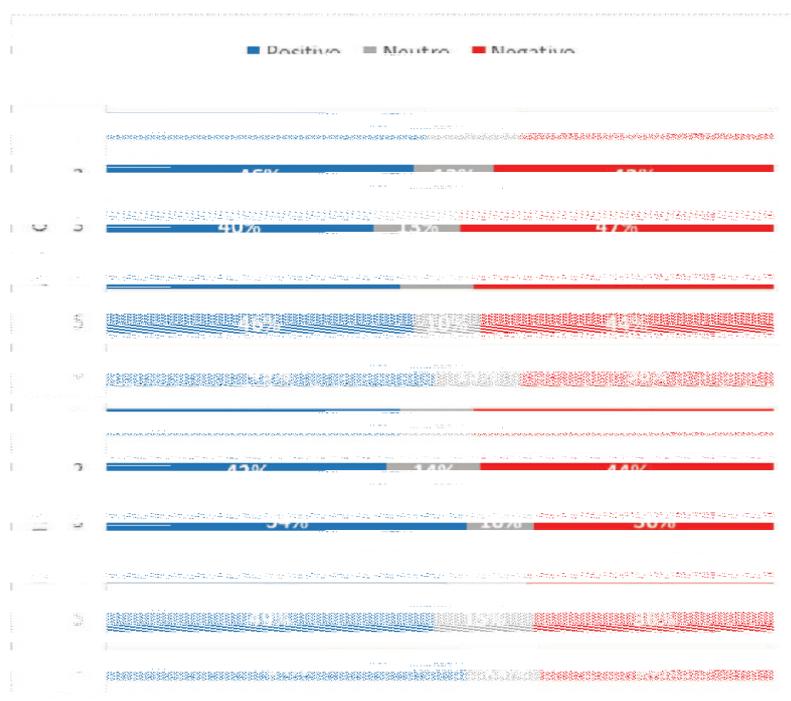


Gráfico 1 Resultados da análise de sentimento aplicada em seis conjuntos de dados contendo *tweets* relacionados aos candidatos a presidente da república nas eleições presidenciais de 2014 no Brasil.

Fonte: Dados da pesquisa.

O Gráfico 1 demonstra que na reta final do segundo turno das eleições presidenciais houve um crescimento de *tweets* positivos para ambos os candidatos. Os números apresentados contradizem a afirmação de Mohammad et al. (2014) de que os *tweets* eleitorais transmitem emoções negativas duas vezes mais que emoções positivas. Neste estudo há até certa vantagem de *tweets* positivos em alguns conjuntos de dados. A candidata Dilma Rousseff, que iniciou o segundo turno com maior frequência de *tweets* negativos, apresentou o maior índice de *tweets* positivos registrado na pesquisa por meio da análise realizada no sexto conjunto de dados. Já, o candidato Aécio Neves, apesar das variações de *tweets* positivos e negativos, durante o período do segundo turno, apresentou resultados muito próximos, se considerarmos apenas a análise do primeiro e do último conjunto de dados. Aqui é importante frisar que o fato de um usuário se manifestar negativamente em relação a um candidato não quer dizer que ele seja favorável ao outro. Dessa forma, o gráfico representa apenas a frequência relativa da ocorrência de *tweets* positivos, negativos e neutros para cada candidato. A Tabela 4 apresenta alguns exemplos de *tweets* classificados nesse estudo.

Quadro 2 Exemplo de *tweets* classificados pela polaridade.

Polaridade	Tweets
Positivo	“Eu voto em Dilma Rousseff meu voto é 13 #Dilma13” (Usuário A, Twitter, 24/10/214).
Neutro	“Dilma e Aécio estão empatados” (Usuário B, Twitter, 14/10/214).
Negativo	“Quem conheceu o desgoverno do #FHC jamais vota em #Aécio #AécioNever” (Usuário C, Twitter, 15/10/214).

Fonte: Elaborado pelo autor.

Como mencionado anteriormente, o intuito deste trabalho é descobrir se a análise de sentimento, baseada em dados extraídos de mídias sociais pode revelar as preferências políticas dos cidadãos como nas pesquisas de opinião tradicionais. Para tanto, foi realizado o confronto dos resultados da análise de sentimento (opiniões positivas e negativas) constantes no Gráfico 1 com os resultados das seis pesquisas eleitorais de intenção de votos e rejeição dos candidatos realizadas pelo Datafolha (2014) durante o segundo turno das eleições presidenciais no Brasil. Primeiramente, foram comparadas a pesquisa de intenção de votos com as opiniões positivas extraídas do Twitter. O Gráfico 2 traz os resultados das seis pesquisas de intenção de votos (DATAFOLHA, 2014), o Gráfico 3 apresenta apenas os *tweets* positivos relacionados aos candidatos Aécio Neves e Dilma Rousseff e o Gráfico 4 confronta os resultados da pesquisa de intenção de votos com os *tweets* positivos de cada candidato separadamente.



Gráfico 2 Resultados das seis pesquisas de intenção de votos para presidente da república no segundo turno.

Fonte: DATAFOLHA, 2014.

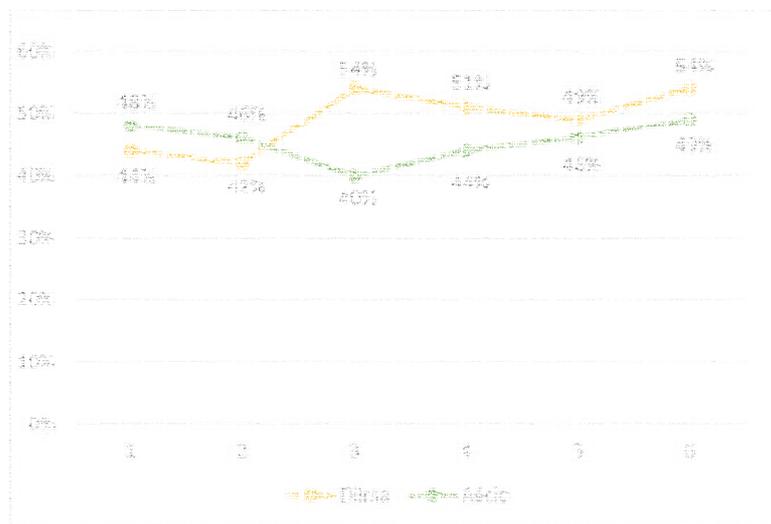


Gráfico 3 Tweets positivos relacionados aos candidatos.
Fonte: Dados da pesquisa.

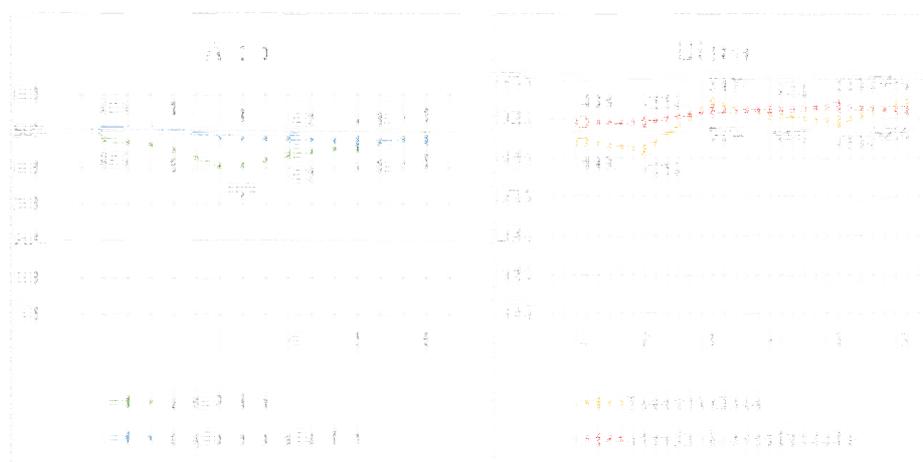


Gráfico 4 Comparação entre a pesquisa Datafolha (2014) de intenção de votos e os *tweets* positivos identificados na análise de sentimento.
Fonte: Dados da pesquisa.

Analisando os Gráficos 1 e 2 separadamente, percebe-se que o volume de *tweets* positivos relacionados à candidata Dilma Rousseff se tornou maior que

os do candidato Aécio Neves, no mesmo momento em que o Datafolha (2014) identificou que a referida candidata havia assumido a liderança das intenções de voto. O Gráfico 4 demonstrou que a frequência de *tweets* positivos dos dois candidatos, durante o segundo turno das eleições, manteve-se próximo aos resultados da pesquisa de intenção de votos, tendo maior aproximação de resultados com diferença de apenas 1% enquanto o resultado mais distante apresentou uma diferença de 8%.

Da mesma forma que foram confrontados os *tweets* positivos com a pesquisa de intenção de votos, os *tweets* negativos identificados, na análise de sentimento, também, foram comparados aos resultados da pesquisa de rejeição dos candidatos do Datafolha (2014). O Gráfico 5 resume os resultados das seis pesquisas de rejeição dos candidatos a presidente da república. O Gráfico 6 apresenta apenas os *tweets* negativos relacionados aos candidatos e o Gráfico 7 confronta os resultados da pesquisa de rejeição com os *tweets* negativos de cada candidato separadamente.

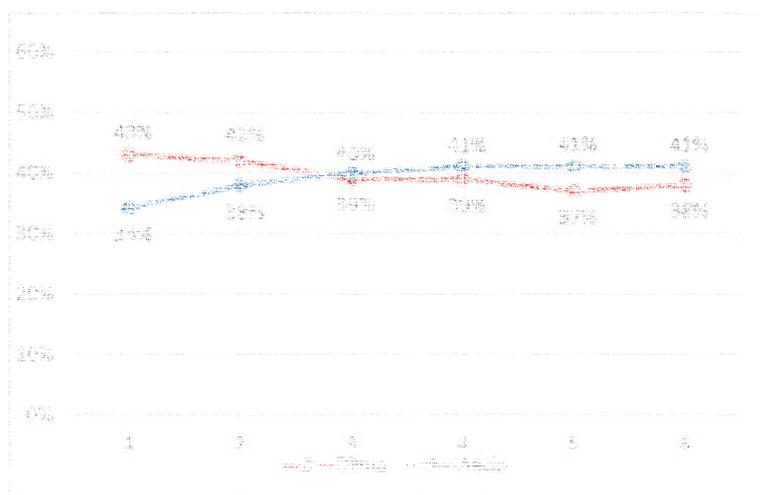


Gráfico 5 Resultados das seis pesquisas de rejeição dos candidatos a presidente da república.

Fonte: DATAFOLHA, 2014.

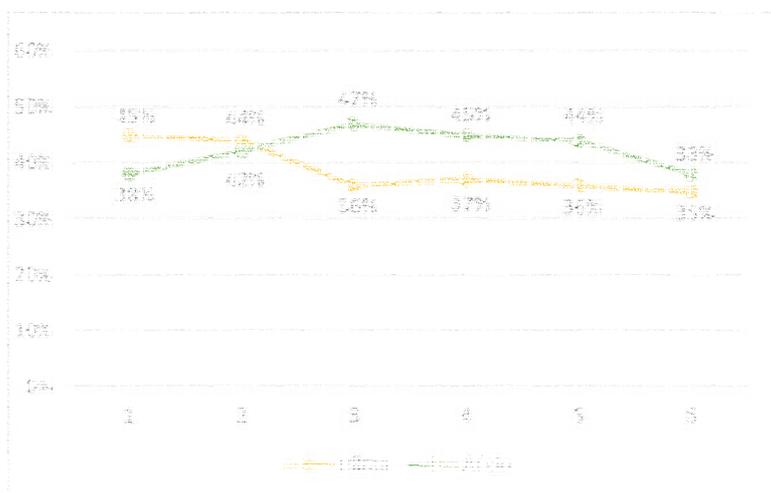


Gráfico 6 Tweets negativos relacionados aos candidatos.
Fonte: Dados da pesquisa.

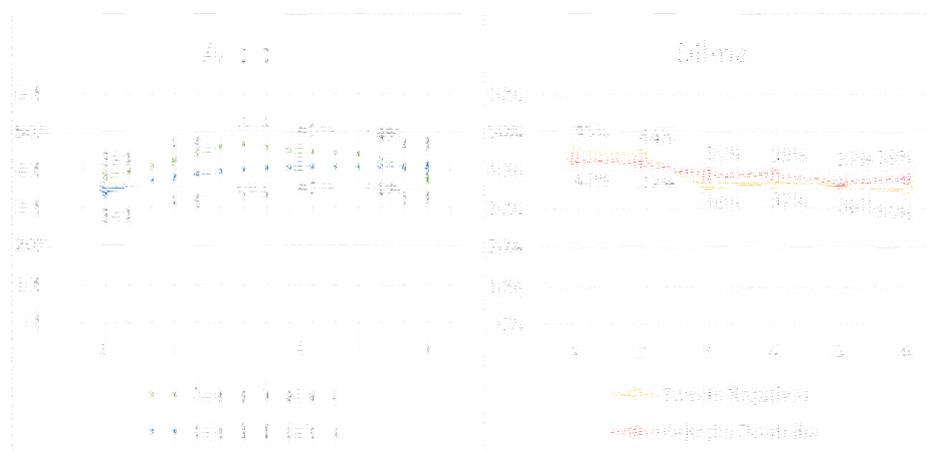


Gráfico 7 Comparação entre a pesquisa Datafolha (2014) de rejeição dos candidatos e os *tweets* negativos identificados na análise de sentimento.
Fonte: Dados da pesquisa.

A comparação demonstrada no Gráfico 7 entre os resultados da pesquisa de rejeição dos candidatos (Gráfico 5) com o volume de *tweets* negativos,

obtidos na análise de sentimento (Gráfico 6), revelou grande proximidade de resultados, principalmente, nas pesquisas relacionadas à candidata Dilma Rousseff, com variação média de 2% em relação à pesquisa Datafolha (2014). Nessa comparação os resultados mais discrepantes pertencem às pesquisas relacionadas ao candidato Aécio Neves que apresentou 7% de diferença entre os resultados na comparação do terceiro conjunto de dados.

8 CONCLUSÃO

Este estudo buscou identificar se os resultados, obtidos pela aplicação da técnica de análise de sentimento em dados extraídos de mídias sociais, pode revelar as preferências políticas dos cidadãos com precisão aproximada das pesquisas de opinião tradicionais. Tendo como objeto de estudo o segundo turno das eleições presidenciais do Brasil em 2014, foi realizada a análise de sentimento, a partir de *tweets* contendo termos-chave relacionados aos candidatos Aécio Neves e Dilma Rousseff. Posteriormente, os resultados dessa análise foram confrontados com os resultados de seis pesquisas de intenção de votos e de rejeição dos candidatos realizada pelo Datafolha (2014).

De forma geral, o confronto dos resultados das pesquisas Datafolha (2014) com os resultados da análise de sentimento apresentou números bastante aproximados. Mesmo que a análise de sentimento não tenha atingido o mesmo nível de precisão da pesquisa de opinião tradicional, a técnica apresentou resultados satisfatórios que apontam a opinião dos eleitores. Desse modo, o fluxo de informações do Twitter pode ser aproveitado de forma significativa para identificar as preferências dos eleitores (TUMASJAN et al., 2011). Embora os usuários do Twitter e de outras mídias sociais não sejam necessariamente representativos de toda a população de cidadãos de um país (CERON et al., 2014), os resultados deste estudo demonstraram que a análise de sentimento

pode indicar a preferência dos eleitores com resultados próximos das pesquisas de opinião tradicionais.

Questões como o baixo custo envolvido na análise de sentimento (O'CONNOR et al., 2010; BACHNER; HILL, 2014), a grande quantidade de dados gratuitos disponíveis nas mídias sociais (CERON et al., 2014; HE; ZHA; LI, 2013) e a importância que essas mídias têm atualmente para a sociedade como uma forma de difusão de opiniões e outras informações (KHAN; BASHIR; QAMAR, 2014; KORENEK; ŠIMKO, 2014; STEENKAMP; HYDE-CLARKE, 2014), fazem com que a análise de sentimento seja uma importante aliada para identificar opiniões (positivas e negativas) dos cidadãos sobre diversos assuntos (CERON et al., 2014; YU; DUAN; CAO, 2013).

Assim, enquanto os métodos de análise de sentimento não atingirem precisão superior aos métodos tradicionais de pesquisa de opinião, a análise automatizada de dados extraídos do Twitter pode servir como um complemento para as pesquisas tradicionais.

Este estudo apresenta algumas limitações. Primeiramente a análise de sentimento e as pesquisas eleitorais do Datafolha (2014) foram realizadas, a partir de dois universos distintos, ou seja, a opinião dos usuários do Twitter não representa, necessariamente, a opinião dos eleitores que não se manifestaram em mídias sociais. Por isso, os resultados da análise de sentimento aqui apresentada não pode representar toda a população de eleitores.

Com a finalidade de superar estas limitações, estudos futuros devem se preocupar em traçar o perfil dos usuários de mídias sociais, identificando estratos sociais que representam melhor o universo de eleitores. Também são necessários estudos que busquem medir até que ponto a discussão nas mídias sociais influencia a opinião pública em favor de um candidato ou partido político. E não menos importante, é preciso procurar por novas técnicas ou

mesmo combinar métodos existentes visando atingir níveis mais altos de precisão e representatividade na análise de sentimento.

REFERÊNCIAS

BACHNER, J.; HILL, K. W. Advances in public opinion and policy attitudes research. **Policy Studies Journal**, Urbana, v. 42, p. S51-S70, May 2014.

BAKLIWAL, A. et al. Sentiment analysis of political tweets: towards an accurate classifier. In: WORKSHOP ON LANGUAGE ANALYSIS IN SOCIAL MEDIA, 2013, Atlanta. **Proceedings...** Atlanta: Association for Computational Linguistics, 2013.

BEYER, Y. Using discovertext for large scale twitter harvesting. **Microform & Digitization Review**, London, v. 41, n. 3-4, p. 121-125, Dec. 2012.

CERON, A. et al. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. **New Media & Society**, Milano, v. 16, n. 2, p. 340-358, Mar. 2014.

CHOY, M. et al. **US presidential election 2012 prediction using census corrected twitter model**. Cornell: Cornell University, 2012. Disponível em: <<http://arxiv.org/ftp/arxiv/papers/1211/1211.0938.pdf>>. Acesso em: 22 jun. 2014.

COCHRAN, W. G. **Sampling techniques**. New York: John Wiley & Sons, 2007.

COLLIS, J.; HUSSEY, R. **Pesquisa em administração: um guia prático para alunos de graduação e pós-graduação**. 2. ed. Porto Alegre: Bookman, 2005.

DATAFOLHA. Pesquisas eleitorais: rejeição dos candidatos. **Instituto de Pesquisas Datafolha**, São Paulo, out. 2014. Disponível em: <<http://eleicoes.uol.com.br/2014/pesquisas-eleitorais/brasil/rejeicao/>>. Acesso em: 28/10/2014.

DISCOVERTEXT. Disponível em: <<http://www.discovertext.com>>. Acesso em: 22 mar. 2014.

EIRINAKI, M.; PISAL, S.; SINGH, J. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, New York, v. 78, n. 4, p. 1175-1184, July 2012.

FORTUNY, E. J. D. et al. Media coverage in times of political crisis: a text mining approach. **Expert Systems with Applications**, New York, v. 39, n. 14, p. 11616-11622, Oct. 2012.

GAYO-AVELLO, D. **I wanted to predict elections with twitter and all i got was this lousy paper**: a balanced survey on election prediction using twitter data. Oviedo: Department of Computer Science, 2012.

GHIASSI, M.; SKINNER, J.; ZIMBRA, D. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. **Expert Systems with Applications**, New York, v. 40, n. 16, p. 6266-6282, Nov. 2013.

GUERRA, P. H. C. et al. From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 17., 2011, San Diego. **Proceedings...** California: ACM, 2011. p. 150-158.

HE, W.; ZHA, S.; LI, L. Social media competitive analysis and text mining: a case study in the pizza industry. **International Journal of Information Management**, Guildford, v. 33, n. 3, p. 464-472, June 2013.

JUNGHERR, A.; JÜRGENS, P.; SCHOEN, H. Why the pirate party won the german election of 2009 or the trouble with predictions: a response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting elections with twitter: what 140 characters reveal about political sentiment". **Social Science Computer Review**, Durham, v. 30, n. 2, p. 229-234, May 2012.

KHAN, F. H.; BASHIR, S.; QAMAR, U. TOM: twitter opinion mining framework using hybrid classification scheme. **Decision Support Systems**, Amsterdam, v. 57, p. 245-257, Jan. 2014.

KHAN, G. F. et al. From e-government to social government: twitter use by Korea's central government. **Online Information Review**, Bradford, v. 38, n. 1, p. 95-113, Jan. 2014.

KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4065-4074, Aug. 2013.

KORENEK, P.; ŠIMKO, M. Sentiment analysis on microblog utilizing appraisal theory. **World Wide Web**, London, v. 17, n. 4, p. 847-867, July 2014.

KUMAR, A.; SEBASTIAN, T. M. Sentiment analysis on twitter. **International Journal of Computer Science Issues**, Oxford, v. 9, n. 3, p. 372-378, July 2012.

LARSSON, A. O.; MOE, H. Studying political microblogging: twitter users in the 2010 swedish election campaign. **New Media & Society**, Milano, v. 14, n. 5, p. 729-747, Aug. 2011.

LIU, B. **Sentiment analysis and subjectivity**. 2. ed. Chicago: Handbook of Natural Language Processing, 2010.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: AGGARWAL, C. C.; ZHAI, C. (Ed.). **Mining text data**. Chicago: Springer, 2012. Cap. 13, p. 415-463.

MANZA, J.; BROOKS, C. How sociology lost public opinion: a genealogy of a missing concept in the study of the political. **Sociological Theory**, San Francisco, v. 30, n. 2, p. 89-113, June 2012.

MOHAMMAD, S. M. et al. Sentiment, emotion, purpose, and style in electoral tweets. **Information Processing & Management**, Elmsford, Oct. 2014. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0306457314000880>>. Acesso em: 12 mar. 2015.

MOSTAFA, M. M. More than words: social networks' text mining for consumer brand sentiments. **Expert Systems with Applications**, New York, v. 40, n. 10, p. 4241-4251, Aug. 2013.

MOUTHAMI, K.; DEVI, K. N.; BHASKARAN, V. M. Sentiment analysis and classification based on textual reviews. In: INTERNATIONAL CONFERENCE ON, 2013, [S.l.]. **Proceedings...** [S.l.]: [Information Communication and Embedded Systems, 2013. p. 271-276.

O'CONNOR, B. et al. From tweets to polls: linking text sentiment to public opinion time series. In: CONFERENCE ON WEBLOGS AND SOCIAL

MEDIA, 2010, Washington. **Proceedings...** Washington: AAAI, 2010. p. 122-129.

O'CALLAGHAN, D. et al. **Online social media in the syria conflict:** encompassing the extremes and the in-Betweens. Cornell: Cornell University, 2014.

OLIVEIRA, D. J. S.; BERMEJO, P. H. S.; SANTOS, P. A. Sentiment analysis, social media, and public administration. in: ĆEMAL, D. et al. (Ed.). **Handbook of research on democratic strategies and citizen-centered E-government services.** Hershey: IGI Global, 2015. Cap. 13, p. 231-249.

PERRIN, A. J.; MCFARLAND, K. Social theory and public opinion. **Annual Review of Sociology**, Palo Alto, v. 37, n. 1, p. 87-107, Aug. 2011.

RAMBOCAS, M.; GAMA, J. **Marketing research:** the role of sentiment analysis. Porto: Universidade do Porto, 2013.

RAPIDMINER. Disponível em: < <https://rapidminer.com>>. Acesso em: 22 fev. 2014.

ROBALDO, L.; DI CARO, L. Opinion mining-ML. **Computer Standards & Interfaces**, Amsterdam, v. 35, n. 5, p. 454-469, Sept. 2013.

SCUP. Disponível em: < <http://www.scup.com/>>. Acesso em: 28 fev. 2014.

SILVA, N. F. F.; HRUSCHKA, E. R.; HRUSCHKA JÚNIOR, E. R. Tweet sentiment analysis with classifier ensembles. **Decision Support Systems**, Amsterdam, v. 66, p. 170-179, Oct. 2014.

SMAILOVIĆ, J. et al. Stream-based active learning for sentiment analysis in the financial domain. **Information Sciences**, v. 285, p. 181-203, Nov. 2014.

SOBKOWICZ, P.; KASCHEKY, M.; BOUCHARD, G. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. **Government Information Quarterly**, Amsterdam, v. 29, n. 4, p. 470-479, Oct. 2012.

SOMASUNDARAN, S.; WIEBE, J.; RUPPENHOFER, J. Discourse level opinion interpretation. In: **INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS**, 22., 2008, Manchester. **Proceedings...** Association for Computational Linguistics, 2008. p. 801-808.

STEENKAMP, M.; HYDE-CLARKE, N. The use of facebook for political commentary in South Africa. **Telematics and Informatics**, Amsterdam, v. 31, n. 1, p. 91-97, Feb. 2014.

TUMASJAN, A. et al. Election forecasts with twitter: how 140 characters reflect the political landscape. **Social Science Computer Review**, Durham, v. 29, n. 4, p. 402-418, Nov. 2011.

VARGO, C. J. et al. Network issue agendas on twitter during the 2012 U.S. presidential election. **Journal of Communication**, New York, v. 64, n. 2, p. 296-316, Mar. 2014.

XU, W. W. et al. Predicting opinion leaders in twitter activism networks: the case of the wisconsin recall election. **American Behavioral Scientist**, Princeton, v. 58, n. 10, p. 1278-1293, Sept. 2014.

YANG, D.-H.; YU, G. A method of feature selection and sentiment similarity for Chinese micro-blogs. **Journal of Information Science**, Cambridge, v. 39, n. 4, p. 429-441, Aug. 2013.

YOON, S.; ELHADAD, N.; BAKKEN, S. A Practical approach for content mining of tweets. **American Journal of Preventive Medicine**, New York, v. 45, n. 1, p. 122-129, July 2013.

YU, Y.; DUAN, W.; CAO, Q. The impact of social and conventional media on firm equity value: a sentiment analysis approach. **Decision Support Systems**, Amsterdam, v. 55, n. 4, p. 919-926, Nov. 2013.

ZAPPAVIGNA, M. Ambient affiliation: a linguistic perspective on twitter. **New Media & Society**, Milano, v. 13, n. 5, p. 788-806, Aug. 2011.

ZHANG, P.; HE, Z. A weakly supervised approach to Chinese sentiment classification using partitioned self-training. **Journal of Information Science**, Cambridge, v. 39, n. 6, p. 815-831, Dec. 2013.

9 CONSIDERAÇÕES FINAIS

Esta dissertação propôs a aplicação da técnica de análise de sentimento em dados extraídos de mídias sociais para identificar a opinião dos cidadãos sobre assuntos relacionados à política e à administração pública, e gerar conhecimentos que possibilitem incorporar a opinião da sociedade civil no corpo político. Sob esse foco foram elaborados três artigos teórico-empíricos com propostas de aplicação da análise de sentimento em situações distintas, mas diretamente relacionadas ao objetivo geral deste trabalho. Enquanto o primeiro artigo buscou identificar se a análise de sentimento pode mensurar a opinião dos cidadãos perante a atuação do governo, o segundo procurou verificar se a técnica pode contribuir para as práticas da Gestão Social estabelecendo uma relação dialógica entre governos e cidadãos e, finalmente, o terceiro artigo testou a eficácia da análise de sentimento em relação a outros métodos de pesquisa tradicionais.

Os resultados dos artigos demonstraram ser possível a utilização da análise de sentimento em mídias sociais para revelar a opinião dos cidadãos, com precisão satisfatória, sobre vários temas, tais como a atuação do governo, programas sociais e eleições.

Diante disso, os artigos contribuíram diretamente para preencher as lacunas descritas na introdução deste trabalho. O primeiro artigo demonstrou que é possível gerar conhecimento, por meio de opiniões extraídas das mídias sociais sobre temas relacionados à política e à administração pública aplicando-se a análise de sentimento. O segundo artigo revelou que a análise de sentimento pode contribuir na coprodução de políticas governamentais convertendo os fluxos comunicacionais das mídias sociais em conhecimento capaz de influenciar as decisões governamentais. Já, o terceiro artigo provou que os resultados gerados pela análise de sentimento se aproximam dos resultados de

pesquisas de opinião realizadas por métodos tradicionais, mesmo apresentando níveis de precisão inferior. Mas, ao ponderar entre as vantagens e limitações da técnica, a análise de sentimento se revela como um bom complemento para as pesquisas de opinião tradicionais.

Algumas limitações dos artigos, também, merecem atenção. Deve haver cautela na interpretação dos resultados aqui apresentados. Principalmente no que diz respeito a generalizá-los, pois o universo aqui explorado é de cidadãos usuários de mídias sociais. Outra limitação é a possível mudança de opinião do indivíduo em relação à percepção social, sentindo-se pressionado a ir de encontro da opinião da maioria. Também faltam estudos que atestem o quanto as discussões realizadas nas mídias sociais representam ou influenciam a opinião dos cidadãos que são alheios ao universo on-line. Por isso, são necessárias futuras pesquisas que se preocupem em explorar novas ferramentas tecnológicas que possibilitem mensurar a opinião dos cidadãos de forma mais precisa e representativa.

ANEXO - Aceite condicional referente ao Artigo 2