

## ***Crowdsourcing e machine learning: uma revisão sistemática com discussão do uso para a participação pública dos cidadãos***

Ricardo Mendes Jr

### **RESUMO**

A cada ano tem aumentado o interesse pelo uso da tecnologia como agente de aumento da participação popular na gestão pública. O uso crescente das mídias sociais como instrumento tanto de movimentos populares quanto da crítica ou apoio de grandes massas a fatos ocorridos ou declarações de figuras públicas pode ser um sinal de que a tecnologia pode ser utilizada também na gestão pública. Este artigo explora a ideia de organizar a inteligência coletiva da população utilizando *crowdsourcing* e *machine learning* na participação pública dos cidadãos. *Crowdsourcing* já é utilizado há algum tempo para coletar dados ou realizar análises que só podem ser feitas por pessoas, incluindo a participação pública em decisões políticas ou de gestão, ao contrário da Inteligência Artificial, que utiliza dados para concluir classificações que podem explicar fatos ou auxiliar as decisões. No entanto, a Inteligência Artificial pode ser associada ao *crowdsourcing*. Assim a pergunta de pesquisa colocada é como melhorar as aplicações de participação pública com *crowdsourcing* com o uso de *machine learning*? Nesta pesquisa são analisadas publicações que apresentam propostas de uso conjunto de *crowdsourcing* e *machine learning*, uma das ferramentas consideradas de Inteligência Artificial. A análise discute o potencial de aplicação destas ferramentas em conjunto, no que foi apresentado nos artigos encontrados, e quais destas aplicações visam a participação pública. As conclusões apresentam o estado atual no tema proposto, a partir desta pesquisa, e as possibilidades de trabalhos futuros.

## INTRODUÇÃO

A política entendida como um *shopping center* está fadada ao fracasso, porque transforma os cidadãos em clientes vorazes e impacientes, e um cliente sempre tem razão. Um cidadão, no entanto, é aquele que se engaja na tomada de decisões coletivas e aceita o resultado destas, ainda que o seu efeito não corresponda às suas preferências particulares. Atualmente temos enfrentado alguns desafios, seja no Brasil ou no restante do mundo, para solucionar os problemas relacionados à gestão pública, seja das cidades ou das nações: como incrementar a qualidade dos bens e serviços públicos? Como equacionar os esquemas de financiamento para as diversas políticas públicas? Como resolver o financiamento na estrutura federativa, com participação do governo federal, estados e municípios? Como aperfeiçoar os instrumentos institucionais e legais para aumentar a eficácia do planejamento, orçamento e gestão públicos? E para introduzir inovações que acompanhem as mudanças que vêm ocorrendo com o avanço estratosférico do uso da tecnologia?

Muitas respostas a estes desafios passam pela mudança na forma de participação popular na gestão pública e na tomada de decisão. Seja pela própria mudança do sistema representativo ou pelo uso massivo da tecnologia. Também já é reconhecido por muitas partes que precisamos gerar mais inteligência nos processos de gestão pública, incorporando as partes interessadas no diálogo. Em todos estes caminhos é necessário o aumento da participação popular, em novas formas, diferentes das atuais, e que possivelmente farão uso da tecnologia.

A cada ano tem aumentado o interesse pelo uso da tecnologia como agente de aumento da participação popular na gestão pública. O uso crescente das mídias sociais como instrumento tanto de movimentos populares quanto da crítica ou apoio de grandes massas a fatos ocorridos ou declarações de figuras públicas pode ser um sinal de que a tecnologia pode ser utilizada também na gestão pública. Assim uma questão que está no contexto deste artigo é: como incluir o cidadão na gestão pública com o uso da tecnologia?

No entanto, a participação popular na gestão pública demanda a construção de uma estrutura que possibilite organizar o processo de gestão da informação de

modo a que de um lado a população conheça os problemas da cidade e as informações completas e de fácil entendimento e de outro lado as informações coletadas da população possam ser utilizadas de forma eficaz pela administração pública. Muitas prefeituras já possuem sistemas de *call center* ou ferramentas para ouvir a população, mas de uma forma ainda que necessita de tratamento ou organização posterior. O que provoca uma perda de eficácia e de tempo para que os resultados possam ser obtidos. Para uso destas plataformas as pessoas precisarão possuir conhecimentos específicos no assunto, e a plataforma precisa ter uma organização interna para depurar e validar as informações. O que exige esforços adicionais além de estruturar o sistema voltados a orientar as pessoas e formar uma equipe de validação, que pode ser voluntária, como é, por exemplo a Wikipédia.

Há também várias aplicações utilizando *smartphones* que já vêm sendo utilizadas para coleta de informação com a população, como o por exemplo, o Colab.re. Muitas outras aplicações existem que não visam diretamente à gestão pública, mas buscam melhorar algum problema da cidade coletando informações das pessoas, como o Waze. Uma característica comum nestas duas aplicações, e importante para a gestão pública, é a geolocalização do dado, que atualmente é muito fácil de ser obtida nos smartphones.

O uso da informação que tem como fonte a população ou multidão é conhecido como *crowdsourcing*. As ferramentas existentes, como as citadas acima, não são totalmente aplicações de *crowdsourcing* por não terem como resultado a busca de um conhecimento sobre determinado problema urbano visando a sua solução. Estas aplicações coletam dados com o objetivo de agregar ou modificar uma informação na sua plataforma. No caso do Colab, a prefeitura tem a sua disposição painel de gestão com os dados coletados. A Wikipédia pode ser considerada o melhor exemplo de aplicação de *crowdsourcing* no sentido que estamos tratando aqui. E adicionalmente, as aplicações existentes utilizam algoritmos de inteligência computacional para organizar as informações na sua plataforma, mas não para extração de conhecimento. O *crowdsourcing* associado à geolocalização é o que é conhecido como mapeamento coletivo. As informações são colocadas pelas pessoas no mapa, seja automaticamente pela

aplicação, como no caso do Colab e do Waze, ou manualmente numa base geolocalizada.

*Machine Learning* é uma classe de técnicas de inteligência computacional que permite extrair informações que podem produzir conhecimento a partir de uma grande quantidade de dados, no caso, os dados coletados por *crowdsourcing*.

Este artigo explora a ideia de organizar a inteligência coletiva da população utilizando *crowdsourcing* e *machine learning*. Inteligência coletiva é uma inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências (LEVY, 2015). A base e o objetivo da inteligência coletiva são o reconhecimento e o enriquecimento mútuos das pessoas, e não o culto de comunidades fetichizadas ou hipostaiadas (LEVY, 2015). Desta forma, observamos que não se pretende apenas coletar dados para alimentar uma base de dados da qual se possa extrair informações, mas mais do que isso – fomentar o surgimento da inteligência coletiva e o uso para a gestão pública, utilizando *machine learning*.

Este artigo apresenta o resultado de uma primeira pesquisa exploratória, com revisão bibliográfica. São analisadas publicações que apresentam propostas ou aplicações de *crowdsourcing* e *machine learning* de forma independente, e também em conjunto, e publicações que apresentem aplicações ou estudos destas técnicas para participação pública. Esta pesquisa limitou-se a artigos publicados ou em publicação em 2018 e 2019.

## **CROWDSOURCING**

*Crowdsourcing* é um fenômeno que tem recebido crescente atenção nas pesquisas acadêmicas e tem ganho cada vez mais aplicações práticas em pesquisas e negócios.

Howe (2006) criou o conceito em 2006 dando a seguinte definição:

*“crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an*

*undefined (and generally large) network of people in the form of an open call”.*

*Crowdsourcing* é considerado um tipo importante de inovação aberta (MARJANOVIC et al., 2012, WIKHAMN e WIKHAMN, 2013) (MOKTER, 2015, p. 3). Consequentemente, *crowdsourcing* é uma das palavras chave mais frequentemente usadas na literatura de inovação aberta (EBNER et al., 2009).

Há algumas revisões sobre *crowdsourcing* na literatura, ESTELLÉS-AROLAS e GONZÁLEZ-LADRÓN-DE-GUEVARA (2012) fizeram um estudo das definições de *crowdsourcing*. Eles encontraram mais de 40 definições em uma extensa literatura. ZHAO e ZHU (2012) fizeram uma revisão sistemática, limitando seu escopo no campo de sistemas de informação. LEIMEISTER et al. (2009).

MOKTER e KAURANEN (2015) fizeram uma revisão ampla apresentando a evolução da literatura em *crowdsourcing*, suas aplicações, benefícios e desafios. Eles encontraram 346 artigos, os quais contêm a palavra “*crowdsourcing*” no título, resumo ou lista de palavras-chave, publicados em inglês, sendo a grande maioria publicados em periódicos não indexados (ISI) e em anais de conferências. Então separaram os artigos para análise em três grupos: artigos indexados (ISI), não indexados e artigos de congressos, considerando que os primeiros são mais científicos por serem revisados pelos pares. Na análise das palavras chave os autores identificaram 997 palavras, sendo as de maior ocorrência: social (50), web (40), innovation (36), open (30), information (24), online (22), community (21), mechanical (20), collective (19) e networks (19). A palavra “mechanical” possivelmente aparece porque a plataforma Mechanical Turk da Amazon é o exemplo mais citado de *crowdsourcing*. “*Collective intelligence*” é um tema bastante discutido na literatura de *crowdsourcing*. E “*community*” frequentemente aparece em conjunto com outras palavras tais como “*based*”, “*detection*”, “*development*”, “*managers*” e “*ranking*” (MOKTER e KAURANEN, 2015, p. 5). Os periódicos com as publicações encontradas pelos autores são de várias áreas, incluindo ciência da informação, geografia, inovação, meio ambiente, pesquisa médica, ciências comportamentais, tecnologia de mídias, serviços e jornalismo. Em 2 periódicos há 2 publicações

em cada, e nos demais há uma única publicação. Em sua análise dos estudos empíricos nos artigos mais citados os autores concluem que *crowdsourcing* pode efetivamente ser útil para projetos de planejamento urbano (BRABHAM, 2009), pode prover informações precisas e em tempo para desastres naturais (GAO et al., 2011), pode ser valiosa para coletar informações geográficas via dados abertos, ferramentas ou serviços (CRAMPTON, 2009), e que amadores, além dos profissionais, podem ser valiosos para coleta de dados geográficos (GOODCHILD, 2009), entre outros estudos. A partir dos artigos pesquisados eles listam as aplicações de *crowdsourcing* existentes: geração de ideias, micro tarefas, *software open source*, participação pública, ciência colaborativa (“*citizen science*”), jornalismo colaborativo (“*citizen journalism*”), wikies (Wikipedia, Geowiki.org, Wikimapia, WikiTerra, Openstreetmap e GLOBE).

### *Crowdsourcing* e Participação Pública

O aumento da participação pública é uma prioridade fundamental no planejamento público, e o *crowdsourcing* é um modelo aplicável para garantir participação em projetos de planejamento público (BRABHAM, 2009; HILGERS e IHL, 2010). Nas palavras de Brabham (2009), o *crowdsourcing* é útil para envolver os cidadãos em projetos de planejamento público para aproveitar seu gênio desconhecido. A participação pública via *crowdsourcing* pode envolver uma ampla gama de pessoas e pode facilitar um diálogo aberto entre os cidadãos e os tomadores de decisão (BUGS et al., 2010; ADAMS, 2011). Muitas organizações governamentais, por exemplo, o governo Obama, fizeram uso extensivo de novas tecnologias de *crowdsourcing* para aumentar a participação pública em várias atividades, como o bem-estar da comunidade (NAM, 2012). Outro exemplo é a Nova Zelândia, onde o parlamento aprovou a “*Wiki Policing Act 2008*”, que permite aos eleitores dialogar com a polícia (HILGERS e IHL, 2010). A lei da Nova Zelândia de 1958 foi apresentada ao público em formato wiki para modificá-la e reescrevê-la. Esta versão wiki da nova lei foi oficialmente aprovada pelo parlamento da Nova Zelândia em 2008. A participação pública via *crowdsourcing* pode ajudar a desenvolver projetos de grande escala com alta precisão, como o Google Earth (FRITZ et al., 2009). Como o *crowdsourcing*

substituí os serviços tradicionalmente construídos por profissionais altamente remunerados, bancos de dados abertos baseados em *crowdsourcing* podem ser criados a um custo muito baixo para o benefício da humanidade (HEIPKE, 2010). SELTZER e MAHMOUDI (2012) argumentam que no planejamento público o *crowdsourcing* é mais efetivo do que a participação cidadã convencional.

### *Crowdsourcing e Machine Learning*

Apesar de todo o avanço conseguido com a Inteligência Artificial, há certas áreas em que um pouco mais de precisão ainda pode ajudar. Tarefas como julgar a condição de um site, validar os detalhes mencionados de uma empresa ou até mesmo estudar os objetos em uma imagem requerem intervenção humana, que, por sua vez, pode ser aprimorada em grande parte usando o *crowdsourcing*. Essa é a razão pela qual os pesquisadores de *machine learning* adotaram o *crowdsourcing* como uma ferramenta há muitos anos para disponibilizar uma grande quantidade de dados para treinar seus programas de forma apropriada. A capacidade da multidão de gerar dados para o conjunto de treinamento de programas de aprendizado de máquina conseguiu manter uma fonte de grande conveniência para os pesquisadores e isso levou a uma descoberta adicional de vários outros algoritmos no gerenciamento das complexidades, aumentando a eficiência e até mesmo depurando os modelos construídos (ABHIGNA, SONI e DIXIT, 2018).

## **MÉTODO**

Este artigo apresenta o resultado de uma primeira pesquisa exploratória, com revisão bibliográfica. São analisadas publicações que apresentam propostas ou aplicações de *crowdsourcing* e *machine learning* de forma independente, e também em conjunto, e especificamente publicações que apresentem aplicações ou estudos de participação pública. Esta pesquisa limitou-se a artigos em periódicos nos anos 2018 e 2019, publicados ou em publicação. Foram

excluídas publicações em congressos e capítulos de livros, e também não foram pesquisadas dissertações e teses. Foram realizadas buscas nas bases Periódicos CAPES, Science Direct, Emerald e Scopus, com as palavras chave “*crowdsourcing*” e “*machine learning*” e “*crowdsourcing*” e “*public/citizen participation*”, no título, resumo ou nas palavras chave. As buscas foram apenas em artigos em língua inglesa. Além do recorte no tempo e tipo de publicação foram filtrados apenas os periódicos das áreas Ciência da Computação (excluídas redes e segurança), Ciência da Informação, Engenharias, “*Decision Sciences*” e Ciências Sociais, sendo excluídas as demais áreas (todas com número menor de publicações, exceto gestão, gestão industrial, estratégia e marketing). Não foram analisadas as quantidades de citações dos artigos. Foram extraídos 139 artigos. Na Tabela 1 são apresentados o número de artigos para os termos de busca e na Tabela 2 para cada uma das bases pesquisadas.

Tabela 1 - Resultados das buscas para os termos de busca

Termos	2018	2019
<i>crowdsourcing</i> e machine learning	91	15
<i>crowdsourcing</i> e public/citizen participation	33	0
Total	124	15

Tabela 2 - Resultados das buscas para as bases

Termos	2018	2019
Periódicos CAPES	7	0
Emerald	3	0
Science Direct	51	15
Scopus	63	0
Total	124	15

Para cada um dos artigos extraídos, foram registrados em uma planilha as seguintes informações: nome do periódico, primeiro autor, título do artigo, ano da publicação, volume, número, páginas, resumo do artigo, autores, palavras chave (dos autores).

Os artigos extraídos estão publicados em 70 periódicos, sendo que 15 destes contém mais de uma publicação, apresentados na Tabela 3 e nos demais 55 periódicos há uma publicação apenas.

Tabela 3 - Periódico com mais de um artigo

Periódico	Número de artigos
Future Generation Computer Systems	18
Procedia Computer Science	13
Expert Systems with Applications	9
Information Sciences	9
Information Processing and Management	8
Knowledge-Based Systems	6
IEEE Access	4
Computer Supported Cooperative Work: CSCW: An International Journal	3
IEEE Transactions on Pattern Analysis and Machine Intelligence	2
Journal of systems and software	2
Machine Learning	2
Semantic Web	2
Urban planning	2

Em relação aos autores, nos 139 artigos extraídos há 431 autores. Em 19 artigos há apenas 1 autor, e o número maior de ocorrência, 45 artigos, têm 2 autores. Seguindo com 28 artigos com 3 autores, 24 artigos com 4 autores e 13 artigos com 5 autores. Apenas 6 artigos têm mais de 5 autores, sendo um artigo com maior número de autores, 18. Em relação às palavras chave indicadas pelos autores, há 289 palavras diferentes. 248 palavras ocorrem apenas 1 vez. As palavras com maior ocorrência estão listadas na Tabela 4, juntamente com as palavras associadas que ocorrem algumas vezes.

Tabela 4 - Palavras chave indicadas pelos autores com maior ocorrência

Palavra	Ocorrências	Palavras associadas
Big data	5	
classification	5	
social media	5	case studies (1), measurement (1)
active learning	6	
crowd	8	annotation (2), clustering (2), computing (2), participation (1), truth (2)
deep learning	9	
natural language	10	processing (8), ambiguity (1) e instruction (1)
machine learning	19	
crowdsourcing	43	projects (2), requirements (2), case studies (1), design (1), promotions (1) e workflow (2).

*Crowd truth* algumas vezes é grafado junto - *crowdtruth*. E *crowded* algumas vezes é utilizado no lugar de *crowd*. Bem como *crowdsourced* no lugar de *crowdsourcing*.

Para a análise dos artigos inicialmente foram lidos os resumos para identificar se o artigo é uma aplicação, um estudo empírico ou artigo de revisão. Os artigos foram ainda separados em *crowdsourcing* exclusivamente, *crowdsourcing* combinado com *machine learning* e *machine learning* exclusivamente. O resultado desta classificação é mostrado na Tabela 5. O único artigo na categoria “Outros” apresenta uma proposta de elaboração ontológica para métodos de aprendizado em ontologias usando uma abordagem baseada em gestão do conhecimento, visando o estudo futuro voltado à web semântica.

Os artigos com *machine learning* foram separados em duas categorias. A categoria “*crowdsourcing* para *machine learning*” utiliza a forma mais tradicional de *crowdsourcing*: uma plataforma ou sistema para coletar os dados das pessoas (*crowd workers* ou cidadãos/*citizen*). E a categoria *machine learning* (exclusivamente) coleta dados diretamente das mídias sociais ou outras fontes (*big data*). No entanto, no senso estrito, ambas podem ser consideradas *crowdsourcing*.

Tabela 5 - Classificação dos artigos pelo tipo de conteúdo

Tipo de conteúdo	2018	2019
<i>crowdsourcing</i> - análise	18	1
<i>crowdsourcing</i> - aplicação	15	2
<i>crowdsourcing</i> - review	5	1
<i>crowdsourcing</i> para <i>machine learning</i> - análise	1	0
<i>crowdsourcing</i> para <i>machine learning</i> - aplicação	51	2
<i>crowdsourcing</i> para <i>machine learning</i> - review	4	0
<i>machine learning</i> - análise	1	2
<i>machine learning</i> - aplicação	26	7
<i>machine learning</i> - aplicação - review	2	0
Outros	1	0
Total	124	15

Uma segunda classificação do conteúdo foi feita procurando identificar a principal aplicação ou tema de estudo dos artigos, apresentada na Tabela 6. Cada artigo foi classificado em apenas uma categoria nesta classificação.

Tabela 6 - Classificação dos artigos pela principal aplicação ou tema de estudo

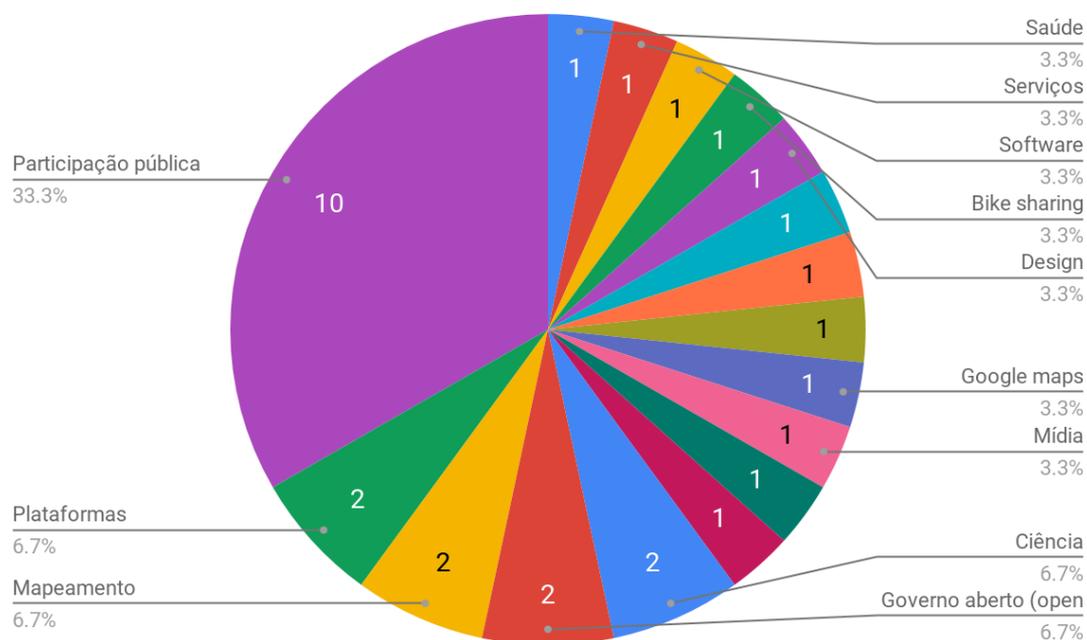
Ocorrências	Aplicação ou tema
1	<i>3D objects; activity recognition; business; chatbots; big data; delivery; face attribute; geo; GLAM (Galleries, libraries, archives and museums); Google maps; industrial; interest; IoV (Internet of Vehicles); learning; machines; map; media (journalism); microelectronics; Neural Network; (crowdsourcing) projects; promotions; reputation; reward; service; value; web; gravitational waves; machine learning.</i>
2	<i>data; quality; (software/engineering) requirements; software; classification bike sharing; design; gov; knowledge ML x NN; Mturk; open gov; security; sentiment; social; astronomy</i>
3	<i>emotions; image; science; social media</i>
4	<i>semantic; mobile; privacy; crowd workers</i>
5	<i>urban; Twitter</i>
6	<i>language; health/medical; IoT</i>
7	<i>active learning; platforms</i>
8	<i>public participation</i>

## RESULTADOS E ANÁLISE

### ***Crowdsourcing***

*Crowdsourcing* é uma ferramenta nova. Nos artigos de revisão já há aplicações relatadas em 2005 (SUROWIECKI, 2005), no entanto o maior número de trabalhos publicados é a partir de 2011 (MOKTER e KAURANEN, 2015). MOKTER e KAURANEN (2015) em seu estudo empírico das aplicações de *crowdsourcing* apontam: geração de ideias, “*microtasking*”, *software* de código aberto, participação pública, ciência cidadã (*citizen science*), jornalismo cidadão (*citizen journalism*) e *wikies*. No presente estudo estão identificadas 17 aplicações e 19 artigos de análise de *crowdsourcing* sendo nos temas indicados no Gráfico 1.

Gráfico 1 - Temas ou aplicações nos artigos de *crowdsourcing*



Os 6 artigos de revisão abordam:

1) as diferentes formas de coletar informações geolocalizadas em *crowdsourcing* (BUBALO et al., 2019);

2) a situação do uso de *crowdsourcing* em galerias, bibliotecas, arquivos públicos e museus ao redor do mundo e sugestões de como a Índia pode se beneficiar desta tendência mundial (GUPTA e SHARMA, 2018);

3) o estado da arte na pesquisa da legibilidade do alfabeto arábico com o objetivo de indicar oportunidades para pesquisas futuras (CAVALLI-SFORZA et al., 2018);

4) analisa a lógica da exploração de uma plataforma pública baseado no retorno social, discutindo a ênfase no balanço entre oferta e demanda como precursor para a estratégia de sustentabilidade (NARDI et al., 2019);

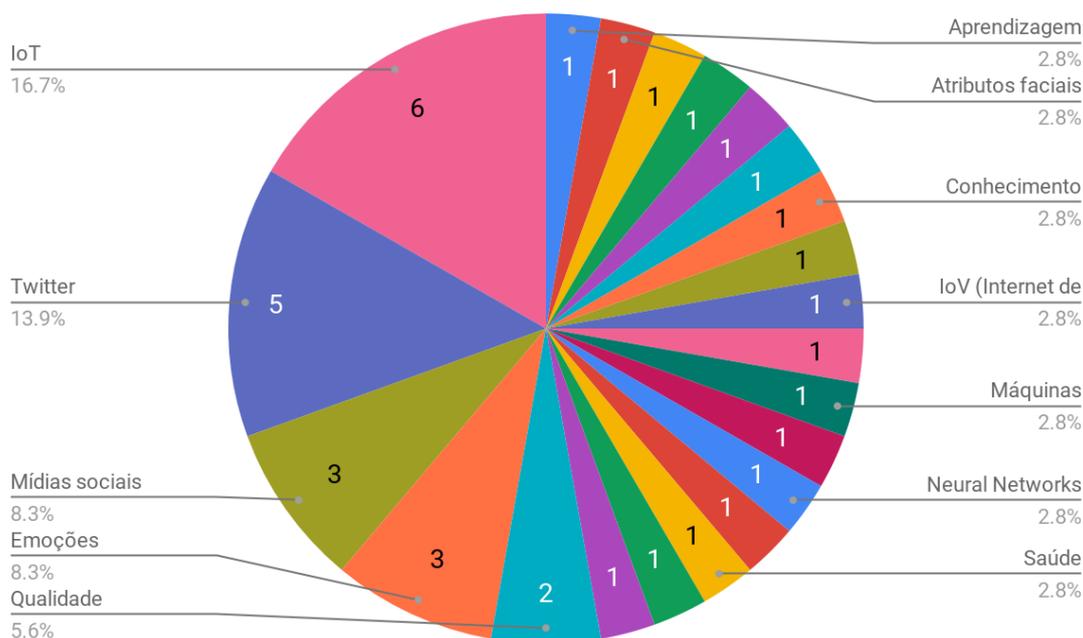
5) revisão sistemática com o objetivo de entender como projetos de *crowdsourcing* são projetados e executados no estado da arte (ASSIS NETO e SANTOS, 2018); e

6) pesquisa o estado da arte da segurança e confiabilidade nas redes de mídias sociais particularmente com o aumento da sofisticação e variedade dos ataques.

## Machine learning

Nos artigos com *machine learning* sem *crowdsourcing* foram encontrados 33 aplicações e 3 análises nos temas indicados no Gráfico 2.

Gráfico 2 - Temas ou aplicações de *machine learning* sem *crowdsourcing*



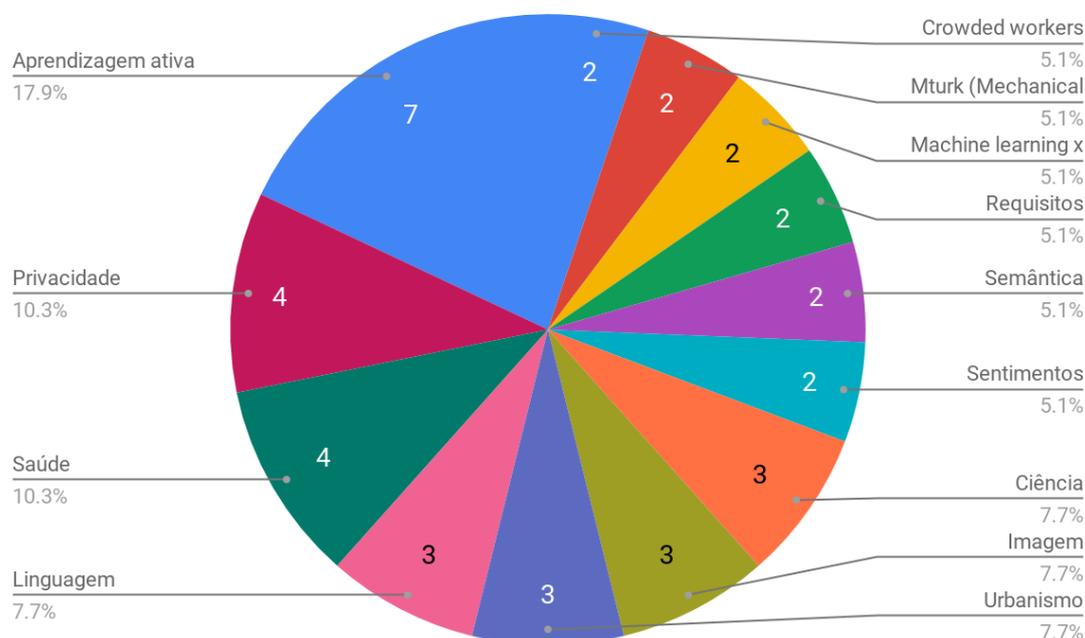
Os 2 artigos de revisão abordam:

- 1) uma sumarização dos métodos Deep Learning para sensores baseados em “mobile” e “wearables” e apresenta alguns desafios para problemas de pesquisa aberta que requerem mais estudos e desenvolvimentos; e
- 2) apresenta uma pesquisa do estado da arte no desenvolvimento de ferramentas de mineração para apoiar a descoberta de “reviews” relevantes para aplicativos “mobile” das lojas.

### Crowdsourcing e Machine learning

Nos artigos categorizados neste estudo como *machine learning* com *crowdsourcing* foram encontradas 53 aplicações com mais de 1 ocorrência e indicadas no Gráfico 3. As aplicações com apenas 1 ocorrência foram: Classificação, Conhecimento, Dados, Design, Emoções, Governo, Indústria, Objetos 3D, Participação pública, Promoções, Recompensas, Reconhecimento de atividade, Senso social (*social sensing*), Valores.

Gráfico 3 - Temas ou aplicações de *machine learning* com *crowdsourcing*



O único artigo encontrado com conteúdo de análises aborda o impacto que o *crowdsourcing* apresenta para as aplicações de *machine learning*. É composto por várias contribuições que *crowdsourcing* pode fazer para melhorar as técnicas que empregam *machine learning* - produzir dados, depuração e verificação de modelos, máquinas inteligentes híbridas para reduzir a intervenção humana necessárias para facilitar o desempenho de alta qualidade pela inteligência artificial e experimentações em desenvolvimento para melhorar interação humano-computador. Apresenta uma discussão sobre a natureza dos trabalhadores da multidão ("*crowd workers*") com ênfase nos vários fatores, como a sua reação a diferentes formas de motivação, seu comportamento uns

em relação aos outros e o dolo entre eles. As conclusões incluem algumas dicas e procedimentos a serem seguidas para alcançar o sucesso através do *crowdsourcing*.

Os 4 artigos de revisão abordam:

- 1) uma análise de aplicações de *crowdsourcing* com *machine learning* utilizando uma metodologia sistemática para classificação, definir características e mostrar o interesse recente nesse tema;
- 2) uma revisão dos trabalhos anteriores dos autores em aplicações para reconhecimento automático da fala utilizando vídeos publicados na web e uma proposta de método para modelar com mais eficiência utilizando aprendizagem não supervisionada;
- 3) uma visão geral de como *crowdsourcing* pode melhorar a pesquisa em *machine learning* e uma revisão da literatura sobre o comportamento dos trabalhadores da multidão (“*crowd workers*”); e
- 4) uma revisão da literatura de processamento de linguagem natural e discussão do ciclo de vida dos dados nas suas aplicações.

## **Participação pública**

Utilizando a classificação em categorias (Tabela 6) podemos indicar 13 artigos que tratam da participação pública, sendo 2 em governo, 1 em governo aberto e 10 em público. Segue um resumo dos artigos que relatam casos práticos de participação popular utilizando *crowdsourcing* (com ou sem o uso de *machine learning*)

Lin (2018) desenvolve um framework para comparar governos inteligentes (*smart govern*) em diferentes contextos institucionais e tecnológicos. Realizam um estudo comparando alguns países do Oeste e a China. Concluem que a governança inteligente está fortemente associada com e-governo e e-democracia em alguns países. Um aspecto comum é que o crescente uso das

mídias sociais, smartphones, portais, plataformas de *crowdsourcing* e sistemas de apoio ao planejamento têm geralmente ajudado a promover os governos inteligentes e os serviços eletrônicos. E isto pode levar a mudanças graduais nas organizações governamentais, a novas relações entre governo, setor privado e cidadãos e melhorias nas cidades.

Shabani e Tairi (2018) explicam a importância do governo aberto e do *crowdsourcing* e examinam o papel que o e-gov representa para conduzir o fenômeno de dados abertos. O estudo abrange os dados abertos disponibilizados em Kosovo.

Komninos (2018) apresenta experiências com a introdução e entrega de dados via *crowdsourcing* e sistemas de compartilhamento de dados, motivados pela urgente necessidade dos cidadãos para resolver um problema com o aparecimento em grande quantidade de medusas na costa oeste da Grécia. Os dados foram produzidos por 13.340 usuários ao longo de 2 meses, relatando cerca 1.800 vezes as condições em 189 pontos na costa. O artigo relata as oportunidades e barreiras para a participação popular em sistemas utilizando os dispositivos móveis.

Clark e Brudney (2018) examinam a representatividade com *crowdsourcing* alcançada nos sistemas de chamada 311 em São Francisco, ao longo dos anos 2011, 2013 e 2015.

Harman e Azzam (2018) desenvolvem um estudo em quatro fases para testar se critérios e padrões podem ser definidos via *crowdsourcing*. O estudo conclui que o *crowdsourcing* tem potencial para ser utilizado na avaliação de juízo de valor para definir critérios e padrões pelo público em geral.

Mitchell e Lim (2018) mostram a complexidade da participação popular em mídias sociais e jornalismo colaborativo utilizando o caso do r/SyrianCivilWar e uma comunidade do Reddit onde usuários discutiram a crise Síria. Demonstram que estas mídias são limitadas na sua capacidade de alcançar a consciência predominante, e questões complexas acabam transformadas em conceitos simples não representativos do contexto e colocadas num grau de importância artificial.

Nummi (2018) analisa a questão se o *crowdsourcing* com a participação popular em um sistema GIS (*geographic information system*) e mídias sociais é um método funcional para revelar valores das pessoas da comunidade, memórias e experiências baseadas nos locais da comunidade, ou seja, o conhecimento da comunidade, para uso em planejamento urbano. Concluem que é possível, mas há vários desafios ao analisar os dados coletados.

## CONCLUSÃO

O presente artigo apresenta uma pesquisa sistemática buscando identificar aplicações e pesquisas unindo *machine learning* e *crowdsourcing*, com ênfase específica nas aplicações para participação pública dos cidadãos.

A pesquisa sistemática foi realizada em artigos publicados em 2018 e 2019, para futuramente ser ampliada para anos anteriores. As palavras chave utilizadas foram “*crowdsourcing*” e “*machine learning*” e “*crowdsourcing*” e “*public/citizen participation*”, no título, resumo ou nas palavras chave. As buscas foram apenas em artigos em língua inglesa. Além do recorte no tempo e tipo de publicação foram filtrados apenas os periódicos das áreas Ciência da Computação (excluídas redes e segurança), Ciência da Informação, Engenharias, *Decision Sciences* e Ciências Sociais, sendo excluídas as demais áreas (todas com número menor de publicações, exceto gestão, gestão industrial, estratégia e *marketing*).

Foram extraídos 139 artigos, sendo 124 artigos de 2018 e 15 de 2019. A maior parte dos artigos resultaram da busca para *crowdsourcing* e *machine learning*, 106. Os artigos extraídos estão publicados em 70 periódicos, sendo que 15 destes contém mais de uma publicação. nos 139 artigos extraídos há 431 autores. Em 19 artigos há apenas 1 autor, e o número maior de ocorrência, 45 artigos, têm 2 autores.

Em relação às palavras chave indicadas pelos autores, há 289 palavras diferentes. 248 palavras ocorrem apenas 1 vez. As palavras com maior ocorrência são: *crowdsourcing* (43), *machine learning* (19), *natural language*

(10), *deep learning* (9), *crowd* (8), *active learning* (6), *social media* (5), *big data* (5) e *i* (5).

Do total de artigos, 17 tratam de aplicações de *crowdsourcing* exclusivamente, 33 de aplicações de *machine learning* exclusivamente e 53 de aplicações combinadas destas duas. Os demais artigos tratam de análises ou review nestes temas. Dos temas tratados nas aplicações ou estudos, participação pública aparece em 10, *active learning* em 7, urbanismo e mídias sociais (Twitter) em 5, semântica, mobile, privacidade e trabalhadores *crowd* (*crowd workers*) em 5. 58 outros temas aparecem em menos de 4 artigos.

Entre as aplicações relatadas nos artigos estão aplicações de *crowdsourcing* na saúde, urbanismo, ciência, mapeamento e governo aberto, entre outras. Nas aplicações de *machine learning* encontram-se saúde, mídias sociais, urbanismo, big data, *bike sharing* e aprendizagem, entre outras. Já nas aplicações conjuntas de *crowdsourcing* e *machine learning* basicamente aparecem os mesmos temas listados acima, acrescentando participação pública e senso social.

Especificamente no tema de participação pública nos 10 artigos encontrados as aplicações e estudos são voltados a coletar informações geolocalizadas, uso em bibliotecas e arquivos públicos, plataformas públicas para definição de estratégias e políticas públicas.

Estes resultados demonstram que a literatura apresenta estudos e aplicações de *crowdsourcing* e *machine learning* em plataformas de participação pública. E metade dos artigos pesquisados já apresentam os dois métodos combinados. O que indica a possibilidade de melhorar as aplicações de *crowdsourcing* com o uso de *machine learning* e a possibilidade concreta de se utilizar estas ferramentas para aumentar a participação pública dos cidadãos.

Para trabalhos futuros propõe-se ampliar a pesquisa para os anos anteriores (até 2015) e elaborar uma análise qualitativa mais aprofundada das aplicações para participação pública na literatura.

## REFERÊNCIAS

ABGHIGNA, B. S., NITASHA, S., DIXIT, S. Dixit, Crowdsourcing – A Step Towards Advanced Machine Learning, **Procedia Computer Science**, Volume 132, 2018, Pages 632-642,

ADAMS, S. A., Sourcing the crowd for health services improvement: the reflexive patient and 'share-your-experience' websites, **Social Science & Medicine**, Vol. 72 No. 7, 2011, pp.

ASSIS NETO, F. R., SANTOS, C. A. S. Understanding crowdsourcing projects: A systematic review of tendencies, workflow, and quality management, **Information Processing & Management**, Volume 54, Issue 4, 2018, Pages 490-506,

BRABHAM, D. C. Crowdsourcing the public participation process for planning projects, **Planning Theory**, Vol. 8 No. 3, 2009, pp. 242-262

BUBALO, M., VAN ZANTEN, B. T., VERBURG, P. H. **Crowdsourcing geo-information on landscape perceptions and preferences: A review**, **Landscape and Urban Planning**, Volume 184, 2019, Pages 101-111.

BUGS, G., GRANELL, C., FONTS, O., HUERTA, J. and PAINHO, M. An assessment of public participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil, **Cities**, Vol. 27 No. 3, 2010, pp. 172-181.

CAVALLI-SFORZA, V., SADDIKI, H., NASSIRI, N. Arabic Readability Research: Current State and Future Directions, **Procedia Computer Science**, Volume 142, 2018, Pages 38-49.

CLARK, B. Y. e BRUDNEY, J. Citizen Representation in City Government-Driven Crowdsourcing. *Computer Supported Cooperative Work*, Online First. 2017.

CRAMPTON, J. W. Cartography: maps 2.0, **Progress in Human Geography**, Vol. 33 No. 1, 2009, pp. 91-100.

ESTELLÉS-AROLAS, E. e GONZÁLEZ-LADRÓN-DE-GUEVARA, F. Towards an integrated crowdsourcing definition, **Journal of Information Science**, Vol. 38, 2012, No. 2, pp. 189-200.

FRITZ, S., McCALLUM I., SCHILL, C., PERGER, C., GRILLMAYER, R., ACHARD, F. e OBERSTEINER, M. Geo-Wiki.org: the use of crowdsourcing to improve global land cover, **Remote Sensing**, Vol. 1 No. 3, 2009, pp. 345-354.

GAO, H., BARBIER, G. and GOOLSBY, R. Harnessing the crowdsourcing power of social media for disaster relief, **IEEE Intelligent Systems**, Vol. 26 No. 3, 2011, pp. 10-14.

GOODCHILD, M. NeoGeography and the nature of geographic expertise, **Journal of Location Based Services**, Vol. 3 No. 2, 2009, pp. 82-96.

GUPTA, D. K., SHARMA, V. Analytical study of crowdsourced GLAM digital repositories, **Library Hi Tech News**, Vol. 35 Issue: 1, 2018, pp.11-17,

HARMAN, E. e AZZAM, T. Incorporating public values into evaluative criteria: Using crowdsourcing to identify criteria and standards. **Evaluation and Program Planning**. V. 71. 2018. pp. 68-82.

HEIPKE, C. Crowdsourcing geospatial data, **ISPRS Journal of Photogrammetry and Remote Sensing**, Vol. 65 No. 6, 2010, pp. 550-557.

HILGERS, D. and IHL, C. Citizensourcing: applying the concept of open innovation to the public sector, **International Journal of Public Participation**, Vol. 4 No. 1, 2010, pp. 67-88

HOWE, J. The rise of crowdsourcing, **Wired** , Vol. 14 No. 6, 2006, pp. 1-4. <https://www.emeraldinsight.com/doi/full/10.1108/SO-12-2014-0029#>

HOSSAIN, Mokter, KAURANEN, Ilkka. Crowdsourcing: a comprehensive literature review, **Strategic Outsourcing: An International Journal**, Vol. 8 Issue: 1, 2015, pp.2-22, <https://doi.org/10.1108/>

KOMNINOS, A. Pro-social behaviour in crowdsourcing systems: Experiences from a field deployment for beach monitoring. **International Journal of Human-Computer Studies**. V. 124. 2019. pp. 93-115.

LEIMEISTER, J.M., HUBER, M., BRETSCHEIDER, U. and KRCCMAR, H. Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition, **Journal of Management Information Systems**, Vol. 26 No. 1, 2009, pp. 197-224.

LÉVY, P. **A inteligência coletiva: por uma antropologia do ciberespaço**. São Paulo: Edições Loyola, 2015, 214 p.

LIN, Yanliu A comparison of selected Western and Chinese smart governance: The application of ICT in governmental management, participation and collaboration. **Telecommunications Policy**, V. 42 (10), 2018. pp. 800-809.

MARJANOVIC, S., FRY, C. e CHATAWAY, J. Crowdsourcing based business models: in search of evidence for innovation 2.0, **Science and Public Policy**, Vol. 39 No. 3, 2012, pp. 318-332

MITCHELL, S. S. D. e LIM, M. Too crowded for crowdsourced journalism: Reddit, portability, and citizen participation in the Syrian crisis. **Canadian Journal Communication**. V. 43 (3), 2018.

NAM T. Suggesting frameworks of citizen-sourcing via Government 2.0, **Government Information Quarterly**, Vol. 29 No. 1, 2012, pp. 12-20.

NARDI, V., ZARPELON, F., VERSCHOORE, J. e ARAÚJO, M., The Multiple Bases of Social Return Platforms: Evidence of a Brazilian Initiative, **International Journal of Public Administration**, Vol. 42, No. 4, 2019, pp. 310-319.

NUMMI, P. Crowdsourcing local knowledge with PPGIS and social media for urban planning to reveal intangible cultural heritage. **Urban Planning**. V. 3 (1), 2018.

SELTZER, E. and MAHMOUDI, D. Citizen participation, open innovation, and crowdsourcing: challenges and opportunities for planning, **Journal of Planning Literature**, Vol. 28 No. 1, 2012, pp. 3-18.

SHABANI, I. e TAIRI, Q. Open Government data in Kosovo and crowdsourcing. **International Journal of Civil Engineering and Technology**. V. 9 (4), 2018. pp. 1333-1343.

STERN, E., GODES, O. and SVORAY, T. Web-based and traditional public participation in comprehensive planning: a comparative study, **Environment and Planning B: Planning & Design**, Vol. 36 No. 6, 2009, pp. 1067-1085.

SUROWIECKI, J., **The Wisdom of Crowds – Why the Many are Smarter than the Few**, New York: First Anchor Books Edition, 2005.

WIKHAMN, B.R. and WIKHAMN, W. Structuring of the open innovation field, **Journal of Technology Management & Innovation**, Vol. 8 No. 3, 2013, pp. 173-185.

ZHAO, Y. and ZHU, Q., Evaluation on crowdsourcing research: current status and future direction, **Information Systems Frontiers**, Vol. 16 No. 3, 2012, pp. 417-434.