

Identificação de Problemas de uma Smart City a partir de posts em redes sociais utilizando algoritmos de machine learning

Proposta de estudo de caso na cidade de Balneário Camboriú

Leonardo Cavalheiro Croda, Daniel Fernando Anderle,
Rodrigo Ramos Nogueira
Instituto Federal Catarinense
IFC
Camboriú, Brasil
lccroda@gmail.com, daniel.anderle@ifc.edu.br,
rodrigo.nogueira@ifc.edu.br

Resumo—Muitas vezes o poder público não leva em conta a opinião da população para tomar decisões que impactam diretamente na mesma, uma possível causa deste problema é que o mesmo não tem formas realmente eficientes de ouvir as demandas dos cidadãos, sendo assim esse trabalho propõe o desenvolvimento de uma aplicação que busca publicações da população local nas redes sociais e utiliza de técnicas de aprendizado de máquina para fazer a análise de sentimento do texto, bem como classificá-lo em um dos pilares das cidades inteligentes.

Palavras-chave—cidades inteligentes; mineração de dados; Data Warehouse; processamento de linguagem natural; redes sociais.

I. INTRODUÇÃO

Uma *Smart City* (Cidade inteligente) é uma cidade que coloca as pessoas no centro do desenvolvimento, incorpora tecnologias da informação e comunicação na gestão urbana. Para isto utiliza esses elementos como ferramentas que estimulam a formação de um governo eficiente, que engloba o planejamento colaborativo e a participação cidadã [1].

É considerada uma *Smart City* uma cidade com bom desempenho em 6 áreas-chave: [economia inteligente, pessoas inteligentes, governança inteligente, mobilidade inteligente, ambiente inteligente e vida inteligente] do desenvolvimento urbano, construída com base na combinação inteligente de doações e atividades de cidadãos confiantes, independentes e conscientes [2].

A cidade de Balneário Camboriú é conhecida como a Dubai brasileira, conta com 135.688 habitantes fixos, mas na temporada esse número chega a 4 milhões de pessoas revezando-se entre os meses de dezembro e fevereiro, o que faz com que a cidade seja mencionada muitas vezes como a capital catarinense do turismo [3]. A prefeitura da cidade possui uma ouvidoria online em seu site, o sistema permite o gerenciamento das demandas da população, por meio dela, é possível fazer denúncias, reclamações, solicitações, sugestões ou elogios. Porém, é necessário contar com a boa vontade do munícipe de abrir uma reclamação, ou

fazer um elogio por exemplo, o que torna dificultoso um mapeamento, uma vez que nem todos têm conhecimento da disponibilidade de tal recurso.

As redes sociais são instrumentos de conversação e troca aberta de informação. São espaços onde as pessoas têm liberdade para falar sobre qualquer assunto, e essa liberdade torna as redes sociais um fator impulsionador no engajamento de diversos movimentos sociais e reivindicações da população em geral [4].

A opinião pública é uma condição de fundo importante, com base na qual a formação de políticas ocorre e depende [5], porém, se feito de forma direta, o processo de consulta a população pode se tornar extremamente caro para o poder público, se fosse feito de maneira presencial, seria necessário alocar recursos especificamente para este fim, se feito de forma direta, mas virtual, iria-se ter que contar com a boa vontade dos cidadãos.

Devido a dificuldade no uso da atual ferramenta e da intensidade do uso de redes sociais para expressar sentimentos em redes sociais, surge a proposta apresentada neste artigo, que visa aproximar as opiniões dos munícipes aos tomadores de decisão. Para isto, usar os dados de postagens das redes sociais para levantar quais são os principais problemas e os principais pontos positivos da cidade, partindo da opinião do ator que mais deveria ser ouvido pelo poder público, o cidadão.

II. TRABALHOS RELACIONADOS

Em seu trabalho, Figueredo et. al [6] descreveram um *framework* focado em processar *tweets*. Este *framework* tem módulos em pipeline, para aquisição e armazenamento de *tweets* de forma robusta, rápida e em tempo real, além de processamento de linguagem natural e análise de sentimento que alimentam um módulo final de análise e visualização.

Em sua tese Anderle [7] ocupou-se em estudar e apresentar um Modelo de Conhecimento baseado em tecnologias como: ontologias, padrões de projeto e análise de sentimento, que pudessem representar semanticamente e de

forma genérica as dimensões de uma *Smart City*, buscando colocar sempre como cerne as pessoas assim como as suas demandas.

Em seu trabalho, Suter [8], desenvolveu um *web crawler* capaz de buscar e baixar *tweets* sobre as eleições presidenciais do Brasil de 2018. Propôs métodos de pré-processamento dos textos e enriqueceu semanticamente os mesmos através de análise de sentimentos, classificando-os como positivos, negativos ou neutros e, por fim, armazenou os textos em um *Data Warehouse* para que consultas voltadas a análise possam ser realizadas de forma mais eficiente.

III. PROCEDIMENTOS METODOLÓGICOS

Quanto à natureza, o presente trabalho se enquadra como uma pesquisa aplicada e tecnológica. Pesquisa aplicada objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos. Envolve verdades e interesses locais [9].

Esta pesquisa também se enquadra como pesquisa tecnológica, pois o produto final é conjunto de arquitetura, software, complementado de um conjunto de dados [10].

O trabalho seguirá uma metodologia desenvolvida em 3 etapas, como pode ser observado na Figura 1.

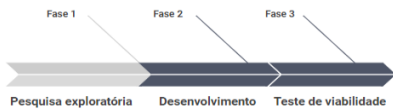


Figura 1. Etapas para conclusão do trabalho

Na primeira etapa foi realizada uma pesquisa exploratória, que tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito [11].

Em um segundo momento a proposta será implementada. Nessa fase primeiramente o *Data Warehouse* será modelado e implementado. *Data Warehouse* é uma cópia de registros informacionais de uma transação especialmente estruturados de forma a que, sobre eles, possam ser elaboradas interrogações e análises [12]. Para armazenar dados em um *Data Warehouse*, é necessário seguir as etapas ETL (extract, transform e load), esta parte do trabalho será desenvolvida como demonstrado na Figura 2.

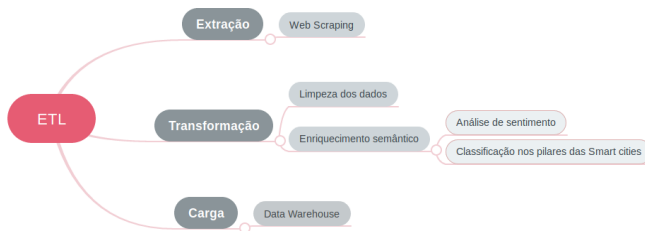


Figura 2. Etapas ETL

Na etapa de extração foi desenvolvido um *web scraper* que recupera informação de posts nas redes sociais sobre Balneário Camboriú, ou que foram postados com localização na cidade, esse *web scraper* tem como base o *web scraper* desenvolvido por [8] em sua etapa de *ETL*. Na etapa de Transformação será realizado o pré-processamento dos textos explorando as técnicas de processamento de linguagem natural existentes. Na etapa de pré-processamento, os textos são transformados da sua estrutura original, para um formato em que possam ser processados de maneira mais simples e eficiente, de acordo com a sua devida finalidade [13].

Após os textos serem pré-processados, serão semanticamente enriquecidos. Enriquecimento semântico trata literalmente de atribuir significados a dados que antes não o possuíam, para isso são utilizadas técnicas de PLN (processamento de linguagem natural), que é uma área de Ciência da Computação que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais [14]. Esse enriquecimento semântico será feito através de análise de sentimento que terá como base o trabalho de Suter [8], classificando as postagens como positivas, negativas ou neutras, além disso, os textos serão classificados de acordo com os pilares das cidades inteligentes, para por fim serem carregados no *Data Warehouse*, o que compreende a etapa de carga.

Com os dados no *Data Warehouse*, serão produzidos dashboards, que terão como principal intuito expor em alto nível os principais problemas da cidade.

Por fim será realizada uma pesquisa com usuários das redes sociais utilizadas no desenvolvimento do trabalho para saber o quão relevantes e realistas as informações levantadas são, assim podendo comprovar se o sistema de fato levanta as demandas da comunidade local.

IV. RESULTADOS ESPERADOS E PARCIAIS

A pesquisa exploratória, já foi concluída e chegou ao seu objetivo de proporcionar maior familiaridade com o tema. Foram executados testes iniciais utilizando o trabalho de Suter [8] como base, sendo que este trabalho compreende as etapas ETL e propõe um *Data Warehouse* parecido com o que será necessário para o desenvolvimento do presente trabalho, porém não se restringindo apenas a posts localizados ou sobre a cidade de Balneário Camboriú e também não realizando a etapa de classificação em um dos pilares das smart cities.

Com a conclusão das próximas etapas espera-se desenvolver uma aplicação que entrega ao poder público informações que antes o mesmo não possuía, e que para isso utilizará de técnicas de aprendizado de máquina e de mineração de dados.

V. CONSIDERAÇÕES FINAIS

Este trabalho apresentou a proposta de um sistema que utiliza de técnicas de aprendizado de máquina e mineração de dados para inferir os principais problemas da cidade de Balneário Camboriú. Para chegar a esse objetivo, é necessário desenvolver as etapas ETL, modelar e implementar um *Data Warehouse* e desenvolver uma aplicação que se alimente dos dados armazenados para construir dashboards que darão uma visão de alto nível dos problemas da cidade que são relatados pelos cidadãos nas redes sociais.

REFERÊNCIAS

- [1] M. Bouskela, M. Casseb, S. Bassi, C. De Luca, e M. Facchina, “Caminho para as smart cities: da gestão tradicional para a cidade inteligente”, Banco Interamericano de Desenvolvimento, 2016.
- [2] R. Giffinger e P.-M. Nataša, Smart cities: Ranking of European medium-sized cities. Centre of Regional Science, Vienna University of Technology, 2007.
- [3] A. P. Neiva, “Balneário Camboriú: os mil encantos da Dubai brasileira”, Diário de Pernambuco, 11-ago-2018. [Online]. Disponível em: <https://www.diariodepernambuco.com.br/noticia/turismo/2018/08/balneario-camboriu-os-mil-encantos-da-dubai-brasileira.html>. [Acessado: 10-jul-2019].
- [4] A. Lemos, “Cidades Inteligentes”, GV Executivo, vol. 12, no 2, p. 46–49, 2013.
- [5] M. Howlett, “A dialética da opinião pública: efeitos recíprocos da política pública e da opinião pública em sociedades democráticas contemporâneas”, Opinião Pública, vol. 6, no 2, p. 167–186, out. 2000.
- [6] M. Figueredo et al., “Using Social Network to Support Smart City Initiatives”, in Proceedings of the 21st Brazilian Symposium on Multimedia and the Web, New York, NY, USA, 2015, p. 101–104.
- [7] D. F. Anderle, “Modelo de Conhecimento para Representação Semântica de Smart Cities com foco nas Pessoas”, Universidade Federal de Santa Catarina, 2017.
- [8] J. V. Suter, “Um *Data Warehouse* baseado no Twitter para análise de sentimento em língua portuguesa: um estudo de caso das eleições de 2018”, Instituto Federal Catarinense, 2018.
- [9] T. E. Gerhardt e D. T. Silveira, Métodos de Pesquisa. PLAGE-DER.
- [10] V. Freitas Junior, C. Woszezenki, D. F. Anderle, R. Speroni, e M. K. Nakayama, “A pesquisa científica e tecnológica”, Espacios, vol. 35, no 9, 2014.
- [11] A. C. Gil, “Como elaborar projetos de pesquisa”, São Paulo, vol. 5, p. 16–17, 2002.
- [12] R. Kimball, The *Data Warehouse* toolkit: practical techniques for building dimensional *Data Warehouses*. John Wiley & Sons, Inc., 1996.
- [13] R. R. Nogueira, . “Newsminer: um sistema de datawarehouse baseado em texto de notícias”, Universidade Federal de São Carlos, 2017.
- [14] R. Vieira e L. Lopes, “Processamento de Linguagem Natural e o tratamento computacional de linguagens científicas”, Em Corpora, p. 183, 2010.