



Análise de redes sociais como apoio na formulação e avaliação de políticas públicas de turismo: o caso do Caminho de Santiago

Analysis of social media for the design and assessment of public tourism policy: the case of the Way of St James

Suso Benitez-Baleato

Equipo de Investigacións Políticas

Universidade de Santiago de Compostela,

jesusmanuel.benitez.baleato@usc.gal

<https://orcid.org/0000-0002-5003-7448>

Susana Sotelo Docío

Grupo Galabra

Universidade de Santiago de Compostela

susana.sotelo.docio@usc.gal

<https://orcid.org/0000-0002-0067-7957>

Recebido/Received: 07/05/2022

Aceitado/Accepted: 05/06/2022

RESUMO:

A gestão de políticas públicas turísticas sustentáveis pode beneficiar-se do conhecimento sobre as percepções tanto de visitantes como de residentes. Se as dos primeiros permitem adaptar melhor a oferta turística às suas demandas, as percepções dos residentes ajudam a identificar possíveis ameaças tanto para a sustentabilidade da atividade turística, como para a avaliação política dos seus gestores. A informação partilhada nas redes sociais demonstrou a sua utilidade para a análise de percepções, pero a metodologia precisa ser adaptada às características específicas do turismo. O objetivo deste artigo é avaliar a utilidade da informação partilhada em redes sociais para o desenho e avaliação das políticas públicas turísticas desde a perspetiva da sustentabilidade para o caso do Caminho de Santiago. Usando algoritmos de inteligência artificial, constatamos uma valoração claramente positiva da experiência atual, e identificamos ações capazes de melhorar a sua sustentabilidade: 1) a preservação e promoção do património natural e cultural, 2) a integração dos visitantes com as comunidades locais, e 3) a melhoria da oferta em áreas como a restauração, a logística ou a segurança. Para facilitar a comprovação da nossa análise, assim como para permitir a reutilização dos nossos dados pela comunidade científica, compartilhamos os resultados através do repositório do Social Data Lab no Dataverse da Universidade de Harvard.

Palavras chave: políticas públicas; redes sociais; Caminho de Santiago; análise de polaridade; processamento da linguagem natural; análise de sentimento.

ABSTRACT:

Sustainable public tourism policy management can benefit from learning about the perceptions of visitors and residents. While the views of the former can help to better satisfy their preferences, the latter can assist in identifying threats to the sustainability of tourism activities, and satisfaction with how tourism is managed. Information shared on social media has proved a useful resource for the study of perceptions, but the approach needs to be adapted in order to deal with the specific features of tourism. The aim of this article is to assess the use of social media in the design and assessment of public tourism policy from the perspective of sustainability and in relation to the specific case of the Way of St James. The artificial intelligence algorithms used show a clearly positive assessment of the tourist experience at present, and identify a set of possible measures to improve the sustainability of the underlying policies: 1) preserve and promote natural and cultural heritage; 2) facilitate integration of visitors with local communities, and 3) improve supply in sectors such as restaurants, logistics and safety. The results of the study have been shared through the the Social Data Lab repository at Harvard University Dataverse in order to facilitate the verification and reuse of the study data by the scientific community at large.

Keywords: public policy; social media; Way of St James; polarity analysis; natural language processing; sentiment analysis.

COMO CITAR ESTE ARTIGO / HOW TO CITE THIS ARTICLE

Benitez-Baleato, Suso e Sotelo Docío, Susana (2022). Análise de redes sociais como apoio na formulação e avaliação de políticas públicas de turismo: o caso do Caminho de Santiago. *Rotur, Revista de Ocio y Turismo*, 16 (2), 56-77. <https://doi.org/10.17979/rotur.2022.16.2.9084>

I. INTRODUÇÃO

O potencial impacto positivo do turismo, tanto a nível económico como social, favoreceu o envolvimento direto das administrações públicas, tornando-se mesmo os principais promotores do fenómeno, como exemplifica a Xunta de Galicia em relação ao Caminho de Santiago. A igual que qualquer outro campo de atuação setorial de uma instituição governamental, as tarefas de promoção e gestão do turismo são realizadas no contexto das *políticas públicas* que identificam e ordenam a sua ação. Os processos de definição, implementação e avaliação dessas políticas públicas requerem uma base empírica suficientemente sólida para garantir a sua adequação ao entorno e a sua evolução. No caso das políticas públicas turísticas resulta de especial relevância a informação sobre as perceções tanto dos visitantes como dos residentes dos lugares de interesse permitindo a deteção precoce tanto de oportunidades como de possíveis ameaças à sustentabilidade da atividade turística. Neste sentido, cabe destacar que a adoção massiva de tecnologias digitais e, concretamente, o uso das plataformas eletrónicas de rede social, favorece a disponibilização de grandes volumes de dados que permitem estudar as perceções sobre o fenómeno turístico. Este trabalho explora a utilidade dos dados partilhados nas redes sociais como base empírica para a avaliação de políticas públicas de promoção turística na área do Caminho de Santiago, desde o ponto de vista da sustentabilidade, utilizando técnicas de análise

computacional baseadas em inteligência artificial que permitem estudar as percepções tanto de visitantes como de moradores.

Os resultados da nossa pesquisa mostram uma percepção claramente positiva da experiência do Caminho, representada por 84% de manifestações favoráveis. Encontramos um claro alinhamento das preferências de caminhantes e moradores em relação ao património natural e cultural, questão que é a principal fonte de percepções expressas tanto positivamente quando estão na viagem, quanto negativamente quando são observadas agressões ou cuidados inadequados. Constatamos um alinhamento semelhante entre visitantes e moradores no referido às culturas locais, especialmente no relativo à gastronomia e às tradições, tanto das localidades situadas diretamente no Caminho, quanto daquelas que estão no seu entorno. Por fim, detetamos uma alta relevância nos aspetos relacionados com a logística da viagem, principalmente nas categorias de assistência à saúde, hospedagem, eventos, transporte, informações meteorológicas, equipamentos desportivos e segurança. Os nossos resultados demonstram a utilidade da abordagem computacional baseada em inteligência artificial para informar as ações de promoção turística no Caminho de Santiago e melhorar a sua sustentabilidade, identificando a centralidade das intervenções voltadas ao cuidado e valorização do património natural e cultural, bem como as oportunidades de promoção local de tradições e culturas e as relacionadas com a logística durante a viagem.

Para apresentar os nossos resultados, primeiramente mostramos o referencial teórico que serve de base para a formulação da nossa hipótese de trabalho. Em segundo lugar, descrevemos a metodologia, com ênfase na abordagem computacional específica utilizada e o uso de algoritmos de inteligência artificial para análise de sentimento em dados extraídos de plataformas eletrónicas de redes sociais, e que exemplificamos com o caso do Twitter. A seguir apresentamos os principais resultados mediante gráficos e tabelas que comentamos em detalhe. Por fim, apresentamos as conclusões do nosso trabalho, enfatizando a utilidade da abordagem computacional para a avaliação e suporte de processos de políticas públicas sustentáveis no campo do turismo.

II. MARCO TEÓRICO

A progressiva participação das instituições públicas na promoção da atividade turística é um fenómeno que se tem registado com especial incidência desde a década de 1970, dando mesmo origem ao desenvolvimento de agências e outras organizações dedicadas especificamente a esta atividade (Pearce, 1992). Como em qualquer outra área setorial, as ações das instituições públicas no campo do turismo ocorrem no contexto das suas *políticas públicas*, definidas de forma genérica como "tudo o que uma organização decide fazer ou não fazer" (Dye, 1992:2; Hall e Jenkins, 1995). Dada a natureza política das instituições públicas, as suas ações estão sujeitas ao escrutínio da cidadania que, em última instância, é quem julga a sua adequação às necessidades do contexto económico e social (Easton, 1965). Para maximizar esta adequação, as políticas públicas são desenvolvidas através de um processo iterativo que, resumidamente, começa com uma fase de *definição* ou *desenho* em que se determinam os objetivos e os meios para os atingir; *implementação*, em que são desenvolvidas as distintas iniciativas; e *avaliação*, onde é determinada a adequação das ações empreendidas, estabelecendo mecanismos corretivos em caso de ser necessário (Birkland, 2015; Gerston, 2014).

Seja na sua fase de desenho, implementação ou avaliação, o sucesso no processo de políticas públicas depende em boa medida da disponibilidade de uma base empírica capaz de informar o analista de forma suficientemente confiável sobre as demandas a serem resolvidas e os meios disponíveis, bem como sobre o desenvolvimento e evolução das ações (Sanderson, 2002). No caso das políticas públicas de turismo, é fundamental dispor de informações sobre as percepções das pessoas envolvidas, que devemos dividir em dois grupos: por um lado, os visitantes, que valorizam a sua experiência em base à sua satisfação, e por outro, os cidadãos que residem nos lugares turísticos, e que valorizam a atividade turística com base no seu impacto económico e social.

Para aceder a estas percepções, resulta de especial interesse a informação que tanto visitantes como residentes partilham através de plataformas de redes sociais eletrónicas, tal como constata os estudos sobre a análise da atividade turística (Zeng e Gerritsen, 2014). O principal incentivo que explica a publicação de informações sobre a experiência turística reside na sua positiva valorização social (Munar e Jacobsen, 2014). No campo empresarial, esta informação é muito valiosa, como ilustra o caso das avaliações emitidas sobre as experiências, e que outros turistas empregam como referência para selecionar recursos turísticos, especialmente na área da restauração e hotelaria (Xiang et al., 2017). O impacto das redes sociais também pode ser visto no uso que os agentes turísticos fazem delas para promover os seus produtos e serviços aos visitantes (Ayeh et al., 2012). Da mesma forma, a informação compartilhada nas redes sociais é muito útil para avaliar as ações governamentais (Singh et al., 2019).

A informação partilhada nas redes sociais inclui um amplo espectro de modalidades, desde informação textual a imagens, vídeos ou lugares geolocalizados, entre outras. Embora seja possível encontrar exemplos de análise de outras modalidades, no campo do turismo a análise de sentimento utilizando a porção textual da informação publicada é de especial interesse, pois expressa, de forma articulada, a experiência partilhada. Esta abordagem tem dado resultados positivos em diversos casos, como a identificação de preferências turísticas para destinos turísticos em Bali (Putra et al., 2019) ou as mudanças derivadas da pandemia de COVID-19 no caso do turismo na Andaluzia (Flores-Ruiz et al., 2021). Relevantemente, o uso de meios digitais foi avaliado em relação à sustentabilidade das atividades turísticas (Ali e Frew, 2012). O nosso trabalho situa-se neste contexto, com o objetivo de explorar a utilidade da análise de dados de redes sociais como suporte para processos de políticas públicas desde a perspetiva da sustentabilidade, tomando como referência o caso do Caminho de Santiago. Para maximizar os benefícios socioeconómicos do Caminho é necessário identificar as novas oportunidades de geração de atividade económica, bem como as possíveis ameaças à sua sustentabilidade devido aos possíveis impactos negativos nas comunidades locais. As ferramentas convencionais de *marketing* continuam a ser úteis, mas não existem metodologias desenhadas para recolher as especificidades do Caminho, como a temporalidade e mobilidade dos peregrinos, a configuração espacial particular dos centros populacionais, ou a sua diversidade linguística e cultural (Mittelstaedt, 2012). Nesse sentido, o conteúdo compartilhado nas redes sociais, por sua espontaneidade e imediatismo, pode contribuir para sanar essas deficiências pela sua superior capacidade para informar sobre as preferências e percepções expressas.

A nossa pesquisa é pioneira nesse sentido, e contribui para o crescente corpus de estudos sobre o impacto socioeconómico do Caminho de Santiago. Exemplos de destaque da crescente atividade nesta área são os projetos de investigação promovidos pela *Cátedra do Camiño de Santiago e das Peregrinacións* da Universidade de Santiago de Compostela, a análise do

impacto do Caminho na cidade de Santiago de Compostela (Torres Feijó, 2020), avaliações económicas (Fernández Méndez et al., 2019), os informes publicados por organismos públicos (OCS 2009, 2017), estatísticas publicadas sobre ocupação hoteleira no Caminho Francês (Xunta de Galicia, 2020) ou relatórios e dados publicados regularmente pelo setor. Embora esses estudos contribuam com achegas valiosas, nenhum considera sistematicamente o fenómeno das redes sociais, nem como objeto de análise, nem do ponto de vista metodológico. Assim, neste trabalho procuramos avaliar em que medida a análise das redes sociais pode servir para identificar oportunidades e ameaças que afetam à sustentabilidade do Caminho de Santiago, contribuindo para a melhoria das políticas públicas relacionadas.

III. METODOLOGIA

Como já vimos, a Internet provou ser uma importante fonte de informação quando se trata de obter dados sobre perceções e experiências, também no campo do turismo. Em particular, redes sociais como Twitter, Facebook ou Instagram são amplamente utilizadas para compartilhar material textual e gráfico que é muito útil ao analisar as perceções em torno de um determinado facto turístico, não só em termos de visitantes, mas também da população local recetora.

O principal objetivo deste trabalho é analisar a polaridade do conjunto de *tweets* relacionados com o Caminho de Santiago, a fim de que possa servir de apoio empírico para a definição de políticas públicas relacionadas com este fenómeno. Nesta secção descreveremos a metodologia utilizada para atingir esse objetivo, que tem três fases: a compilação e armazenamento do corpus de trabalho, a identificação da linguagem dos textos e, finalmente, a análise de polaridade.

Compilação de um corpus de *tweets*

Para a compilação de um corpus de *tweets* utilizou-se uma ferramenta de extração de dados (*web scraper*) adaptada à rede Twitter¹. Uma lista inicial de *hashtags* foi usada como ponto de partida sobre a qual foram aplicados vários processos iterativos de descarregamento e avaliação de amostras até à obtenção da lista final (tabela 1), que serviu como especificação dos termos de pesquisa para obter o corpus.

Foram usados *hashtags* referidos ao Caminho em vários idiomas e em múltiplas variantes, a fim de maximizar a cobertura e assim tentar descarregar o maior número possível de *tweets* relevantes, independentemente de outros fatores, tais como a data de emissão ou a língua em que está escrito.

Por outro lado, a ferramenta de descarga não se limita a recuperar *tweets* que contenham um ou mais dos *hashtags* especificados, mas também extrai aqueles *tweets* que foram reproduzidos (*quoted tweets*) ou incluídos em fios. Isto produz uma certa quantidade de ruído no resultado, sendo resolvido numa fase de pós-processamento para filtrar os *tweets* que não se ajustam aos critérios de pesquisa especificados. Como resultado desse processo de definição de critérios e obtenção e filtragem dos dados, compilámos um total de 415.772 *tweets* relacionados com o Caminho de Santiago, abrangendo o período entre abril de 2009 e

¹ O sistema foi desenvolvido pelo Social Data Lab (<https://socialdatalab.org>) no nó da Universidade de Santiago de Compostela e utiliza chamadas à biblioteca *scrapy* de Python para obter *tweets* através de consultas à interface web, o que permite resolver as limitações da API de Twitter.

junho de 2021. Finalmente, o conjunto de dados extraído foi armazenado numa base de dados relacional (RDBMS).

Tabela 1. Lista de *hashtags* usada para descarregar *tweets*

#caminhodesantiago	#caminoportugues
#caminoasantiago	#caminhoportugues
#caminodelnorte	#xacobeo
#caminodesantiago	#xacobeo2021
#caminofrances	#xacobeo2022
#caminoprimitivo	#viajacobi
#chemindecampostelle	#xacobeo21_22
#jakobsweg	#chemindestjacques
#santiagoway	#germancamino
#stjamesway	#camminodisantiago
#wayofstjames	#chemindesaintjacques
#buencamino	

Fonte: Elaboração própria

Identificação de idioma

A identificação de idioma é uma tarefa do Processamento de Linguagem Natural (NLP) que consiste em determinar automaticamente a língua em que um texto é escrito. É tradicionalmente considerada uma tarefa resolvida (McNamee, 2005) para documentos longos escritos numa única língua. Contudo, textos provenientes de sistemas de *microblogging*, como Twitter, apresentam dificuldades adicionais, tais como textos curtos, registo informal ou mais descuidado, elementos específicos do sistema (como *#hashtags* ou *@mentions*) integrados no próprio discurso ou a combinação de mais de uma língua.

No nosso caso, a identificação da língua é uma parte fundamental do pré-processamento do corpus, não só pela informação em si, mas sobretudo porque a precisão destes dados é importante para garantir o bom funcionamento das fases posteriores do processamento (análise de sentimento ou de polaridade). Cada *tweet* inclui nos seus metadados o código de idioma que indica em que língua está escrito, conforme à classificação automática atribuída pelo sistema². Para determinar a qualidade deste sistema de deteção, selecionámos uma amostra aleatória de 1486 *tweets* em que realizámos a identificação manual da língua. Os resultados são apresentados na tabela 2.

Como se pode ver, o sistema de Twitter apresenta uma boa precisão (92.73%), com apenas um 7.27% de resultados errados. Portanto, para o processamento do corpus decidimos utilizar a classificação de idioma incluída nos metadados em vez de treinarmos um novo sistema, uma vez que consideramos que a sua qualidade é suficiente para este caso e o domínio de trabalho não requer um modelo específico mais adaptado às suas características. Por outro lado, como esperado, uma elevada percentagem dos erros corresponde a pares de

² Não há dados públicos sobre o algoritmo usado pelo reconhecedor de idioma de Twitter. Contudo, um conjunto de dados para a sua avaliação é descrito em Mitjia T (2015).

línguas muito próximas, tais como espanhol-galego-português (60,19%) ou catalão-espanhol (8,33%).

Tabela 2. Avaliação manual da identificação de língua de Twitter para o corpus de *tweets* do Caminho de Santiago

Corretos	1378	92.73%
Não corretos:		
es-gl	39	
pt-gl	17	
pt-es	9	
outros	43	
total incorretos	108	7.27%
Total	1486	100.00%

Fonte: Elaboração própria

Análise de sentimento

No âmbito deste trabalho tomamos como ponto de partida a utilidade da análise das perceções expressas através de plataformas de *microblogging*, como Twitter. A figura 1 mostra dois exemplos: à direita, uma perceção positiva motivada pela beleza do lugar; à esquerda, uma perceção negativa relacionada com o uso de máscaras de proteção. No entanto, esta análise não é possível em todo o corpus devido à sua dimensão. Por este motivo, utilizámos sistemas de deteção automática de polaridade, que facilitam a escalabilidade do corpus e possibilitam que este tipo de análise seja aplicado a grandes conjuntos de dados.

Figura 1. Exemplos de *tweets* com experiências positivas (direita) e negativas (esquerda)



Fonte: Elaboração própria

A análise de sentimento (*Sentiment Analysis*, também conhecida como *Opinion Mining*) é uma técnica de Processamento de Linguagem Natural que permite a identificação das emoções e outros tipos de informação subjetiva num documento, sendo uma das suas aplicações fundamentais a de determinar automaticamente a polaridade positiva, negativa ou neutra de um texto. Esta é uma área com uma longa história em NLP, adaptada a múltiplas línguas, aplicações e algoritmos (Medhat et al., 2014), que é frequentemente aplicada na indústria, especialmente nos setores comercial, de serviços e de atendimento ao cliente.

Em termos gerais, existem duas abordagens principais em sistemas de análise de sentimento: as que fazem uso da aprendizagem automática (*machine learning*) e as que se baseiam em dicionários. Embora as duas tenham vantagens e desvantagens, os métodos de aprendizagem automática supervisionada (como SVM ou Naive Bayes) têm tido tradicionalmente um melhor desempenho (Zhang et al., 2014). Mais recentemente, técnicas de aprendizagem profunda (*deep learning*) também foram aplicadas (Zhang et al., 2018) com resultados que alcançam o estado da arte.

No nosso trabalho utilizámos o kit de ferramentas Linguakit (Gamallo e García, 2017), uma suíte de módulos para diferentes tarefas de NLP, incluindo um módulo de análise de sentimento. O núcleo deste módulo é um sistema híbrido que combina um classificador bayesiano, treinado com textos previamente anotados, e um léxico de polaridade e regras sintáticas para a identificação de marcadores linguísticos que podem mudar a polaridade das palavras (Gamallo e García, op. cit.: 23). Para cada texto fornecido, o analisador de sentimento retorna um valor de polaridade (positivo, negativo ou neutro) junto com a probabilidade.

De todo o corpus, apenas as mensagens em espanhol, português e inglês foram processadas com o analisador de sentimento do Linguakit, resultando num conjunto de dados final de 354.655 mensagens (85,30% do total) contendo a polaridade e a sua probabilidade. Este corpus foi publicado em aberto (Sotelo e Benitez-Baleato, 2021) no repositório do *Social Data Lab* empregando o *Dataverse*³.

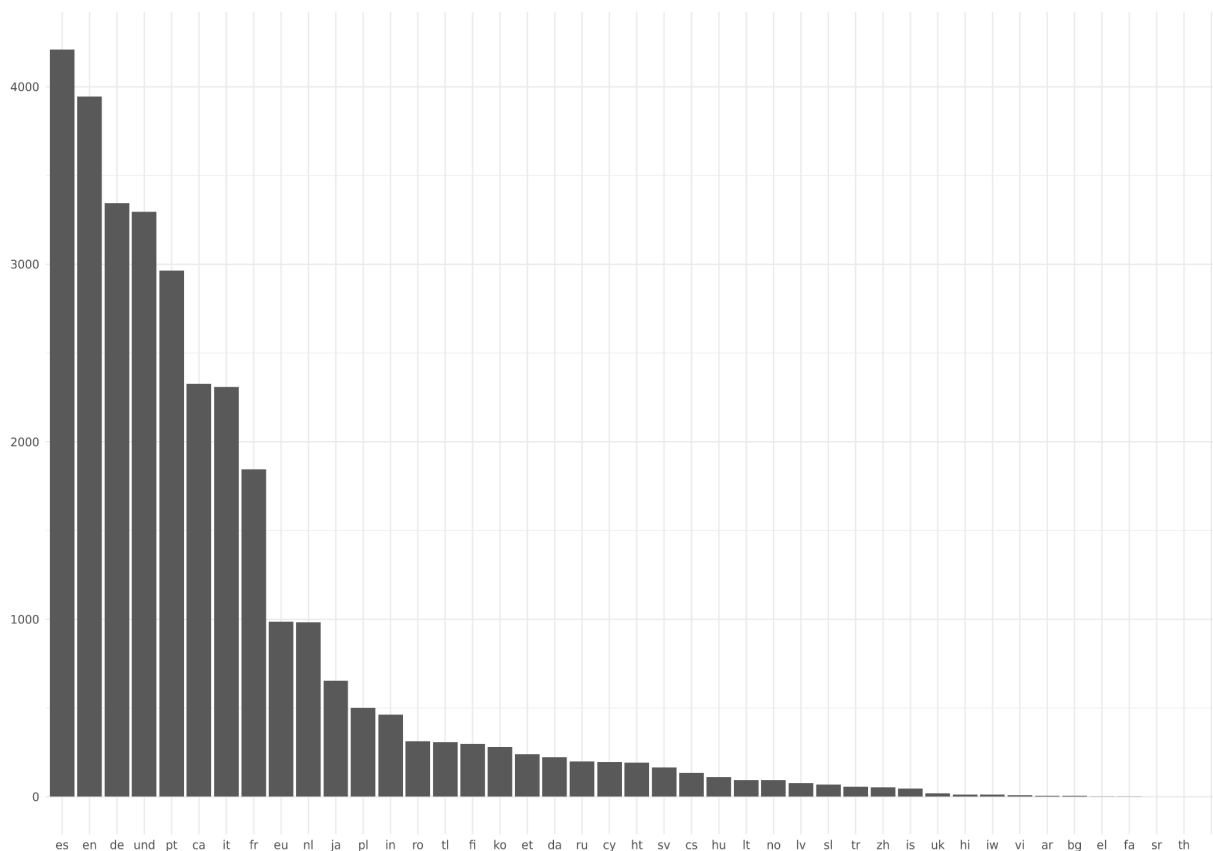
IV. RESULTADOS

Para mostrar os resultados da análise linguística, primeiro procedemos a classificar as mensagens capturadas conforme o idioma utilizado. A figura 2 mostra o peso absoluto de cada idioma para a amostra total.

A informação partilhada está expressa num total de 42 línguas, com uma clara predominância do espanhol (“es”) e do inglês (“en”), seguindo-se um segundo grupo constituído, nesta ordem, por alemão (“de”), português (“pt”), catalão (“ca”), italiano (“it”) e francês (“fr”). A estes dois grupos principais segue-se um terceiro grupo em que se destacam o holandês (“nl”), o japonês (“jp”), o polaco (“pl”) e o basco (“eu”), que se completa com as restantes línguas, entre as quais se destacam o romano (“ro”) e o finlandês (“fi”).

³ <https://dataverse.harvard.edu/>

Figura 2. Peso absoluto de cada idioma

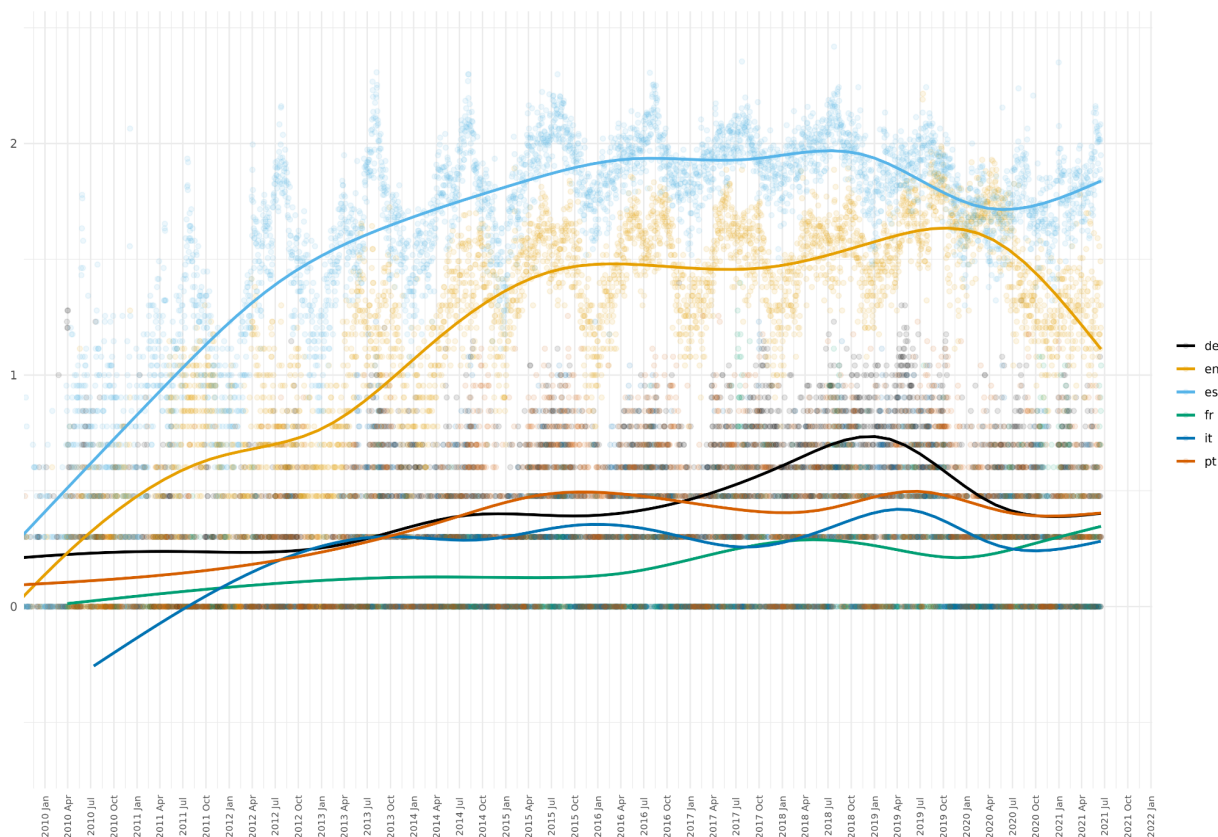


Fonte: Elaboração própria

Como nota técnica, é relevante mencionar que a distribuição absoluta de idiomas também mostra um grupo, o quarto maior do total, que não possui idioma atribuído (marcado 'und' ou 'in' no gráfico). Este grupo representa mensagens para as quais o sistema não conseguiu associar um idioma com certeza suficiente. Para determinar as causas, realizamos um exame de uma amostra aleatória das mensagens afetadas, no qual constatamos que não era razoável atribuir-lhes um idioma, pois eram principalmente mensagens compostas principalmente por imagens, ligações para páginas da web e referências a nomes de lugares, sem texto significativo que permitisse atribuir um idioma de forma confiável. Por esta razão, as mensagens sem atribuição de idioma serão guardadas para uso apenas no caso de análise de palavras-chave, reservando-se as análises baseadas em componentes linguísticos para aquelas com marca de língua e excluindo as indeterminadas.

Após segmentar as informações capturadas pela língua, examinamos o comportamento de diferentes comunidades linguísticas ao longo do tempo para entender a sua importância relativa e a evolução temporal. A figura 3 mostra os resultados dessa análise para os principais idiomas. Para cada instante, o gráfico mostra a tendência esperada calculada pela regressão linear, registrando as mudanças de tendência antes da sua ocorrência. Apresentamos os dados em escala logarítmica para facilitar a visualização, pois a grande distância entre o primeiro grupo e o segundo dificultaria a interpretação correta do último.

Figura 3. Distinto peso dos idiomas



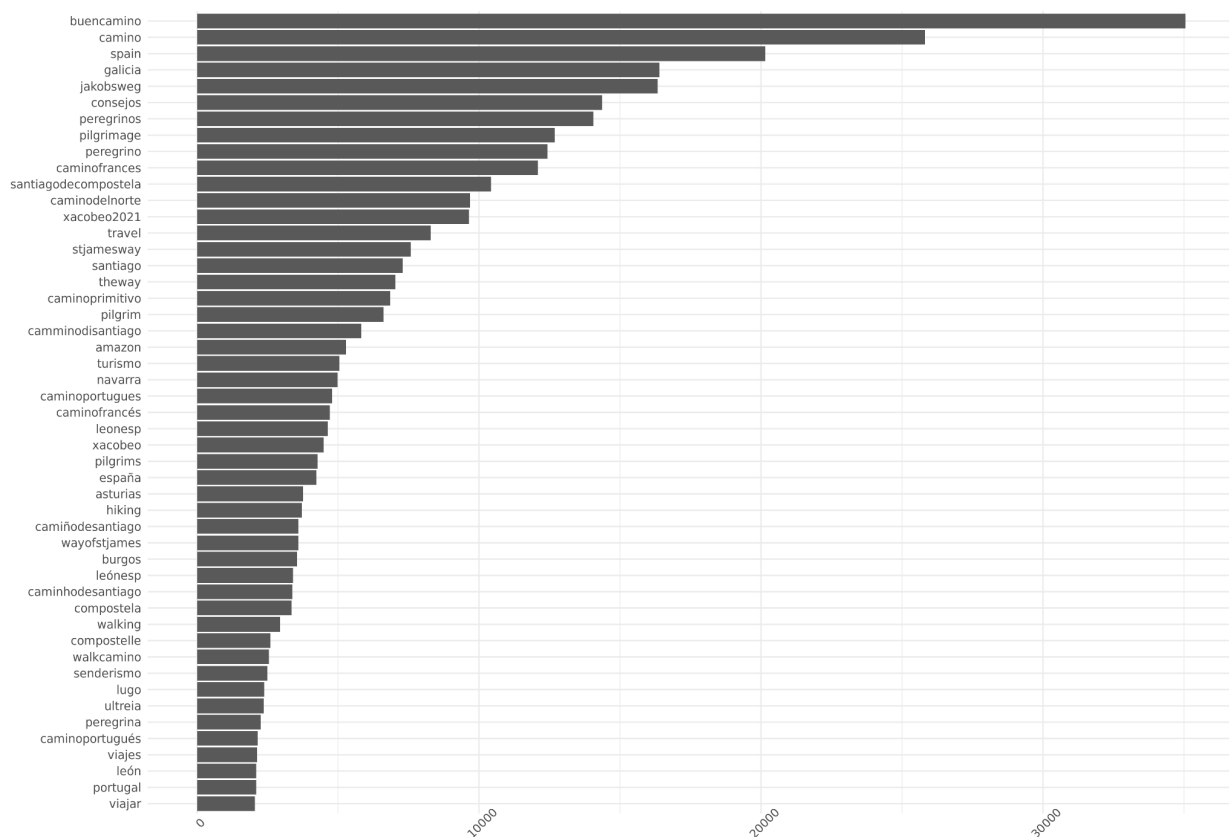
Fonte: Elaboração própria

Os dados confirmam a prevalência detetada no nível agregado, que também se encontra na desagregação temporal. Observa-se a predominância absoluta do espanhol e do inglês, seguido de um segundo grupo (alemão, português, francês e italiano) que apresenta maior variabilidade na prevalência temporal. Assim, enquanto o espanhol continua a ser o idioma que mais é utilizado seguido do inglês, no segundo grupo pode-se observar uma alternância na prevalência de italiano e francês, ao tempo que se constata a perda de protagonismo do alemão sobre o português. No caso do inglês, devido ao seu *status* de língua franca, foi interessante avaliar até que ponto o seu uso corresponde ao de visitantes procedentes de países anglófonos, ou que utilizaram o inglês nas suas comunicações digitais pese a ter idioma nativo distinto. A partir de uma amostra aleatória, pudemos observar a presença significativa de mensagens transmitidas em inglês por falantes de alemão e francês, o que sugere a necessidade de avaliar com mais detalhe a prevalência relativa do inglês para evitar generalizações errôneas sobre essas populações.

Uma vez avaliada a dinâmica idiomática em nível agregado e temporal, passamos a identificar os principais *tokens* ou elementos de relevância linguística das expressões digitais do Caminho de Santiago. Esses *tokens* são muito úteis para análise, pois atuam como índices de conteúdo compartilhado por meio das redes sociais, o que facilita a identificação e captura. Pela natureza da plataforma eletrônica escolhida, o elemento de análise mais destacado corresponde aos chamados *hashtags*, referentes a palavras, ou conjunto delas, utilizadas para rotular ou classificar tematicamente o conteúdo das comunicações.

pelos organismos setoriais. Finalmente, a estratégia de seleção de *hashtags* demonstrou ser capaz de capturar mensagens referidas ao Caminho, deixando fora da mostra aquelas não relacionadas com o fenómeno estudado. Em conjunto, a análise dos resultados no nível temporal, idiomático e de *token* acreditam a capacidade da mostra para ser utilizada como base sobre a que estudar as percepções do Caminho, cujos resultados detalhamos na seguinte secção.

Figura 5. Predominância de *hashtags* excluído o principal

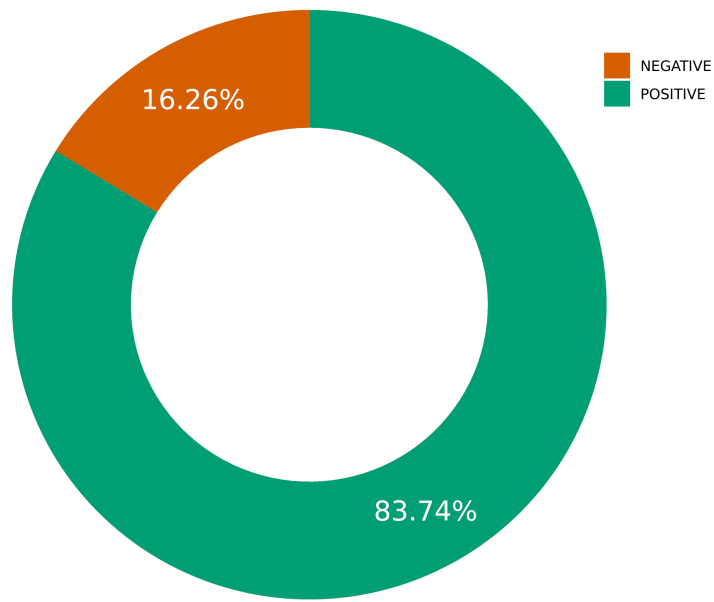


Fonte: Elaboração própria

Análise de percepção

Para realizar a análise de percepções aplicamos os algoritmos de processamento de linguagem natural indicados na metodologia, que nos permitem classificar automaticamente as mensagens conforme a percepção positiva ou negativa que é transmitida. O nosso algoritmo realiza uma classificação em três valores possíveis: percepções positivas, negativas e neutras, que são aquelas que não podem ser enquadradas em nenhuma das outras duas. Assim como fizemos no caso da classificação linguística, examinamos a categoria de mensagens neutras para entender o seu significado e possível impacto na amostra. Como no caso anterior, esta análise revela que as mensagens neutras são compostas principalmente por *links*, imagens e nomes de lugares, além de *tokens* já identificados. Portanto, foram removidas da análise devido à sua inadequação em fornecer informações significativas para a nossa análise. Como pode ser observado na figura 6, aparece uma clara prevalência de mensagens positivas, correspondendo a 84% das mensagens analisadas.

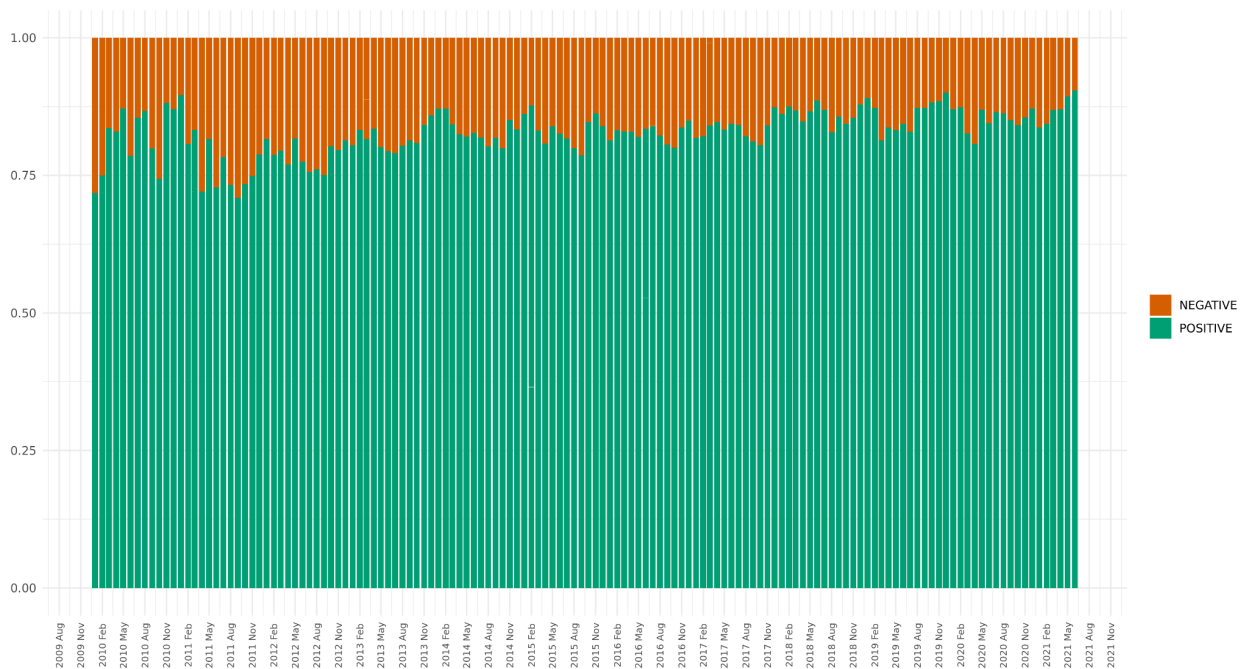
Figura 6. Predominância absoluta de percepções



Fonte: Elaboração própria

Os resultados positivos em relação à percepção do Caminho também são confirmados quando avaliados ao longo do tempo. A figura 7 mostra, para cada mês, a proporção entre percepções positivas e negativas desde 2010, o que confirma a maioria das percepções positivas. Constatase um aumento progressivo da avaliação positiva ao longo do tempo, que se observa com mais intensidade no final dos períodos de férias. Especial atenção deve ser dada ao crescente impacto positivo da avaliação no fim da amostra.

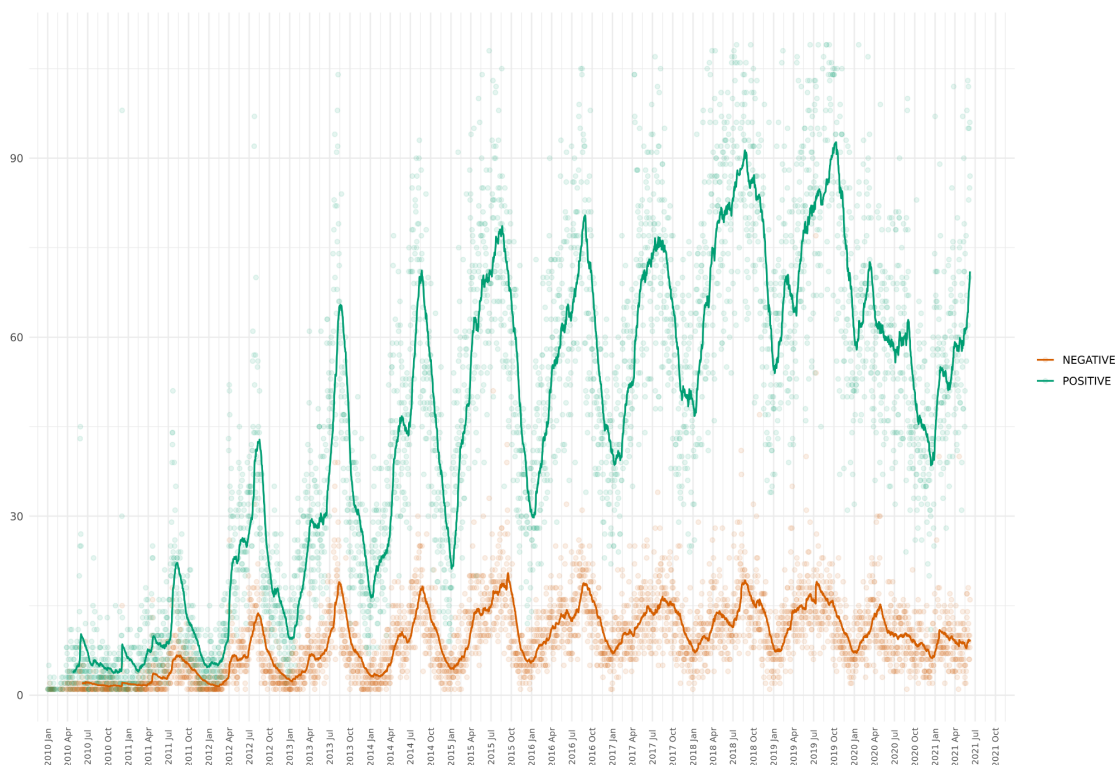
Figura 7. Predominância cronológica de percepções (proporção)



Fonte: Elaboração própria

A figura 8 mostra a mesma informação, agora em valor absoluto para permitir uma visão geral da evolução temporal das perceções desagregadas conforme a sua perceção positiva ou negativa, desde 2010. Este gráfico mostra a evolução temporal do contraste entre perceções negativas e positivas, o que expõe resultados consistentes com análises anteriores e fornece informações mais precisas. Em primeiro lugar, verifica-se a predominância das perceções positivas, representando as negativas apenas 7% no agregado médio do total em valor absoluto, perdendo também peso relativo ao longo do tempo devido ao aumento constante da perceção positiva.

Figura 8. Predominância cronológica de perceções (valor absoluto)



Fonte: Elaboração própria

Análise de oportunidades e ameaças

Verificada a representatividade da nossa amostra e a sua utilidade para reportar sobre a composição linguística, os principais tópicos e as perceções tanto ao nível agregado como na sua desagregação temporal, complementamos a nossa análise com uma avaliação do conteúdo com o objetivo de identificar uma taxonomia de ameaças e oportunidades, bem como um *ranking* de produtos e serviços que possam indicar possíveis demandas insatisfeitas capazes de abrir oportunidades económicas e sociais. Com esse propósito, realizamos iterativamente amostras aleatórias das mensagens captadas nos dois principais grupos linguísticos, identificando os temas mais recorrentes e refinando a categorização até atingir a saturação teórica (Hennink e Kaiser, 2019). Devido à natureza interpretativa desta avaliação, limitamo-nos nesta secção a destacar as conclusões mais relevantes, disponibilizando todos os dados analisados no repositório do Social Data Lab (Sotelo e Benitez-Baleato, 2021) para facilitar a sua comparação por parte da comunidade científica.

O primeiro que pode ser observado é uma clara divisão entre duas fontes de informação: por um lado, os dados são gerados por pessoas que fazem o Caminho em diferentes condições, incluindo turismo cultural e desportivo, ou experiência espiritual e peregrinação religiosa; por outro lado, existe um volume significativo de dados gerados por pessoas ou entidades que residem nos locais por onde passa o Caminho de Santiago ou que geram recursos, como residentes, instituições ou empresas. Essa primeira classificação é consistente com o referencial teórico, que em diversos estudos aponta a importância para a análise das políticas públicas de turismo das percepções tanto dos visitantes dos atrativos turísticos quanto dos moradores e comunidades circundantes. Esta classificação permite-nos discernir temas, percepções e interesses, que se revelam distintos e até dicotómicos, como a análise posterior nos permite determinar. Para facilitar a sua identificação, chamaremos essas duas fontes de 'visitantes' e 'residentes', respetivamente.

A tabela 3 resume os achados classificados como ameaças e oportunidades nos dois grupos indicados. Relativamente à categoria de visitantes, é fundamental notar que a principal fonte de percepções positivas se concentra na fruição da natureza, seguida de um interesse ativo pelo património arquitetónico e cultural. De longe, as expressões de alegria mais frequentes, independentemente da escala temporal, idiomática ou contextual, referem-se à descoberta de lugares naturais, paisagens e recantos em que a natureza - árvores e vegetação em geral, montanhas, pedras, caminhos...- repetidamente obrigam ao caminhante a parar, fotografar o local e partilhá-lo com expressões de alegria e prazer. Da mesma forma, os achados patrimoniais, tanto arquitetónicos como tradicionais, provocam expressões positivas, muitas vezes em relação ao evento de chegada ao destino, especialmente no caso de lugares com elementos de interesse patrimonial. Neste sentido, a chegada a Santiago de Compostela, a entrada no Monte Gozo e a vista para a Catedral concentram as expressões de alegria, que se complementam ao longo do Caminho com as que ocorrem no final da etapa, e muito especialmente na chegada ao albergue. Por fim, outra fonte de satisfação e gozo recorrente é aquela que diz respeito à interação social com outras pessoas ao longo do Caminho, principalmente outros caminhantes, e ocasionalmente em relação ao contacto com moradores, quando há oportunidade de estabelecer uma troca.

Tabela 3. Taxonomia de oportunidades e ameaças

	Oportunidades	Ameaças
Visitantes	Desfrute da natureza e do património	Lesões, fadiga, ruído / civilidade
	Descoberta Gastronómica	Meteorologia, falta de preparação
	Encontros pessoais	Sinalização / orientação
	Chegada a Compostela e pousadas	Segurança: Roubos, Fatalidades
Residentes	Gastronomia e hospitalidade	Agressões ao meio ambiente
	Património e conservação da natureza	Degradação do património cultural
	Aumento da atividade económica	Gestão de recursos turísticos
	Visibilidade, atividade de lazer	Impacto: ruído, lixo, congestionamento

Fonte: Elaboração própria

As expressões associadas às percepções negativas são muito limitadas em volume no caso dos visitantes, em contraste com aquelas originadas nos moradores. No que diz respeito aos caminhantes, são fundamentalmente expressões do sofrimento associado à atividade de andar, fundamentalmente relacionado ao aparecimento de cansaço e feridas no início da experiência. Embora o algoritmo classifique essas expressões como percepção negativa, é necessário esclarecer a interpretação, pois são expressões retóricas que não expressam tanto uma avaliação negativa da experiência quanto dúvidas sobre a própria capacidade de completar o caminho projetado. Identificamos também percepções negativas nas muitas mensagens referentes a 'sofrimento', 'dificuldades', 'fadiga', 'exaustão' que, mais uma vez, são usadas com muita frequência para expressar o desafio de chegar ao destino. Essas expressões muitas vezes aparecem em contraste com o alcance dos objetivos, portanto, a sua consideração como uma percepção negativa não pode se limitar à avaliação algorítmica, recomendando a análise qualitativa para evitar algum viés devido à automação.

Em contrapartida, as expressões negativas registadas nas informações da categoria residente referem-se muito mais intensamente às avaliações negativas. Não se trata tanto de atitudes negativas face ao fenómeno turístico, mas de avaliações de aspetos centrados na percepção das agressões ao ambiente natural e patrimonial das localidades por onde passa o Caminho, bem como ao impacto na gestão dos recursos turísticos. Estas expressões, muitas vezes, traduzem-se em avaliações claramente negativas e críticas, muitas vezes referido a cargos políticos, o que implica o muito que os residentes valoram os seus elementos naturais e patrimoniais, bem como dos recursos turísticos associados, que são percebidos como próprios.

Tal como no caso dos visitantes, avaliamos as percepções negativas dos residentes de forma a detetar neles oportunidades para a promoção de uma atividade turística sustentável, identificando áreas prioritárias de ação e melhoria úteis para instituições e agentes privados. Este é também o caso das percepções positivas dos moradores, muitas vezes referindo-se à visibilidade das próprias localidades e elementos patrimoniais que derivam da sua inclusão em guias e circuitos turísticos, ou eventos relacionados ao Caminho.

A análise de oportunidades e ameaças leva a uma síntese que serve para identificar uma priorização, ou classificação, de produtos e serviços, que dividimos em duas categorias (tabela 4). A primeira categoria, que chamamos de 'património', reflete o alinhamento mais direto entre as preferências expressas pelos caminhantes, de forma dicotómica - positivo primeiro, negativo por último- mas referindo-se aos mesmos elementos. Na mesma linha estão as oportunidades relacionadas com a procura de oferta gastronómica, com uma clara percepção positiva dos caminhantes para a oferta local e especializada que os residentes oferecem nos seus respetivos locais, sobretudo quando associada à cultura e tradições locais. Esta primeira categoria concentra os três grupos de produtos ou serviços com oportunidades de contribuir para uma atividade económica sustentável no Caminho: natureza, restauração e cultura.

No que diz respeito à natureza, ela é a fonte primária tanto de expressões positivas dos caminhantes quando se encontram paragens e paisagens de especial beleza no caminho, quanto de preocupação quando os moradores percebem agressões ou falta de mantimento do entorno natural. É interessante notar o interesse ativo dos caminhantes pela vida animal, seja em relação aos mamíferos domésticos, como cães e gatos, como relacionados às atividades pecuárias, e até mesmo na interação com insetos, todas elas fontes de um grande número de mensagens.

Tabela 4. Ranking de produtos e serviços

Categoria		Produto ou serviço
1	Natureza	Preservação e melhoria do ambiente natural
2	Restauração	Gastronomia 'enxebre' ou cozinha típica local
3	Cultura	Dados do património local
		Circuitos culturais para localidades próximas
		Promover a interação com os moradores e tradições
4	Saúde	Proximidade, imediatismo, transporte
5	Alojamento	Disponibilidade: informações, alternativas locais
		Silêncio: tampões de ouvidos; área de armário separada
6	Eventos	Promoção específica, maior distribuição nas localidades
7	Logística	Transporte e armazenamento de bagagem e bicicletas
8	Meteorologia	Previsão de adversidades e assistência
9	Equipamento	Material de substituição: calçado e vestuário
10	Segurança	Cartão de carteira pré-pago, vigilância

Fonte: Elaboração própria

O segundo elemento do *ranking* tenta refletir as oportunidades no campo gastronómico, especialmente a gastronomia tradicional ou típica de cada localidade, sendo valorizada pelos caminhantes de forma extraordinária. Bebidas típicas e produtos como pães ou queijos e cozinha tradicional são especialmente valorizados e procurados. O elemento cultural que é detetado na gastronomia, também é observado no terceiro lugar do *ranking*, onde se concentram os produtos e serviços relacionados à atividade cultural. Neste sentido, a informação específica sobre os elementos patrimoniais das próprias localidades é valorizada de forma muito positiva, expressando mais recentemente o desejo de conhecer as das localidades próximas, talvez sob a forma de circuitos culturais organizados que os caminhantes possam aproveitar.

Observamos um certo contraste entre o Caminho entendido como desporto e o Caminho entendido como lazer, em que não é tão importante conhecer as etapas no tempo, mas descobrir coisas novas, explorar as áreas próximas, e desfrutar de novas experiências. Neste sentido, destaca-se o interesse ou fruição dos elementos relacionados com elementos espirituais e artísticos ligados às origens ancestrais das respetivas culturas, entre os quais se destacam as referências celtas e germânicas no património da Galiza, como cruces, cruzeiros ou igrejas, bem como os de componente religioso.

Os demais elementos do *ranking* são sequenciados num segundo grupo útil para identificar produtos e serviços mais diretamente ligados ao que chamamos de 'trânsito' ao longo do Caminho. Estes incluem todos os relacionados com elementos logísticos, com especial

destaque para os cuidados de saúde associados a pequenas lesões e desgaste da atividade de caminhada, bem como todos os relacionados com a qualidade do bem-estar e descanso. Nesse sentido, a proximidade e imediatismo dos serviços de saúde ou dos locais de alojamento estão entre os elementos que registam maior protagonismo. São recorrentes as expressões relativas às dificuldades para dormir, principalmente aquelas referentes ao ruído noturno, com duas fontes: os sons naturais emitidos durante o sono pelos demais peregrinos, e aqueles derivados de comportamento incivil, especialmente associado ao consumo de bebidas alcoólicas ou as causadas pela preparação do dia quando estas ocorrem nas primeiras horas da manhã no mesmo quarto em que dormem outros caminhantes. Há também interesse pelas questões logísticas, especialmente no transporte e armazenamento de bagagens e bicicletas, bem como tudo relacionado com a disponibilidade de alojamento e outros recursos. Detetam-se também oportunidades para melhorar as experiências relacionadas com a espera e as filas, causa comum de frequentes manifestações de frustração. Neste grupo identificamos também as demandas relacionadas com a meteorologia, tanto em termos de disponibilização de informações precisas com suficiente antecedência, quanto na resolução de adversidades que ocorrem ao longo do caminho, sendo muito comuns as referências às dificuldades causadas pela chuva em relação a calçados e roupas inadequados ou que precisam ser substituídos ou restaurados.

Por fim, embora com menor incidência, observamos um elemento relacionado à segurança, devido à necessidade de compartilhar espaços em situação de vulnerabilidade com estranhos. Aqui é possível observar uma potencial utilidade nos produtos e serviços capazes de ajudar a minimizar riscos, como cartões de pagamento e telefonia pré-carregada, bem como a presença de Forças e Corpos de Segurança ao longo do percurso, tanto em termos de vigilância dissuasória, como especialmente em relação ao cuidado de pessoas com dificuldades, geralmente associadas a dificuldades de saúde. As notícias de roubos, assaltos e mortes, embora escassas, têm grande impacto quando ocorrem, especialmente em comunidades linguísticas não peninsulares.

V. CONCLUSÕES

O principal objetivo do nosso estudo foi avaliar em que medida a análise das redes sociais pode ser útil para apoiar os processos de políticas públicas de turismo no caso do Caminho de Santiago, numa perspetiva de sustentabilidade. Para tal, desenvolvemos uma metodologia computacional que nos permitiu captar um volume significativo de mensagens partilhadas por visitantes e residentes através das redes sociais, permitindo análises quantitativas e qualitativas. Os resultados obtidos permitiram melhorar o conhecimento sobre a estruturação linguística dos visitantes, quantificar a perceção positiva maioritária e identificar uma série de ações possíveis capazes de acrescentar a sustentabilidade da atividade turística ligada ao Caminho de Santiago, bem como a identificação de possíveis produtos e serviços capazes de aumentar a satisfação com a experiência. Na totalidade, o nosso trabalho confirma a hipótese sobre a utilidade das redes sociais para apoiar os processos de definição, implementação e avaliação das políticas públicas de turismo.

Em primeiro lugar, a nossa análise confirmou a representatividade da amostra capturada, que coincide com a dinâmica cíclica da atividade turística, que aumenta nos períodos de verão, em contraste com a menor atividade da temporada baixa. A análise dos idiomas empregados permitiu identificar os grupos linguísticos maioritários, que podemos dividir em dois grupos: por um lado, o espanhol e o inglês, com clara predominância do primeiro com um acentuado declínio do segundo no final da amostra; e outro grupo composto por alemão,

francês, italiano e português, que apresenta maior variabilidade na prevalência, com aumento acentuado no português. Essa análise permite caracterizar com mais precisão a configuração cultural dos visitantes, o que também é útil para a implantação de ações de comunicação que podem ser mais eficazes ao permitir uma melhor segmentação de públicos. No caso do inglês, a nossa análise sugere a conveniência de uma avaliação mais detalhada da sua utilização como língua franca pelas comunidades alemã e francesa, devido à sua propensão a se expressar também em inglês. Igualmente seria preciso melhorar a identificação de língua fornecida por Twitter, para superar as limitações detetadas no referido às línguas peninsulares, e que afetam de forma específica ao idioma galego e ao catalão.

O nosso estudo permitiu também avaliar a percepção das mensagens, captadas em termos positivos ou negativos. A análise de sentimento mostra uma percepção claramente positiva que é observada em 84% das mensagens, e que aumenta no final da amostra, o que também se constata na desagregação temporal. A análise das reações negativas possibilitou identificar a sua origem maioritária nas comunidades locais, que surgem como resposta ante casos de agressão ou falta de cuidado com o património cultural e natural. Este resultado foi possível ao complementar com análise qualitativa o tratamento automatizado fornecido pelos algoritmos de inteligência artificial, sugerindo ademais a necessidade de melhorar a deteção de sentimentos em casos como as expressões de sofrimento próprias da atividade de caminhar, que não necessariamente implicam uma avaliação negativa das políticas públicas turísticas. É neste ponto que o nosso trabalho realiza a sua achega mais valiosa, pois permite destacar um claro alinhamento nos interesses de visitantes e moradores, sendo justamente o património natural e cultural a causa principal de expressões positivas no caso dos visitantes, e de negativas nos residentes. Acreditamos que esta é uma constatação importante que pode ajudar a melhorar a sustentabilidade das políticas públicas de turismo, sendo que as ações de promoção e melhora do património cultural e natural permitiriam satisfazer as preocupações dos residentes ao tempo que incrementaria a satisfação dos visitantes. Esse mesmo alinhamento entre moradores e caminhantes pode observar-se em termos de culturas locais, especialmente em gastronomia e tradições, tanto dos lugares localizados diretamente no Caminho quanto dos situadas no seu entorno. Conjuntamente, esses achados fornecem uma base sólida para a formulação de políticas públicas em termos de sustentabilidade, pois o foco na proteção e promoção do património natural e cultural resultaria benéfica tanto para visitantes como para residentes. Dessa forma, ações de órgãos públicos orientadas à proteção e melhoria do ambiente natural, de edificações ou paragens singulares, ou de aproximação dos visitantes às comunidades locais por meio da gastronomia e tradições, atuariam para melhorar a experiência do visitante, enquanto fortaleceriam o bem-estar das comunidades locais. Além disso, a análise detalhada das mensagens captadas permitiu-nos identificar um conjunto de áreas de oportunidade para a melhoria dos produtos e serviços exigidos pelos visitantes. Neste sentido podemos destacar os relacionados com a melhoria da logística, quer na preparação da viagem, como no trânsito durante a caminhada, bem como os relacionados com a garantia do bem-estar, a segurança e o desfrute da experiência.

Em resumo, o nosso estudo mostra a utilidade da análise de redes sociais para apoiar a formulação de políticas públicas sustentáveis de turismo no caso do Caminho de Santiago. A nossa abordagem computacional demonstrou a capacidade necessária para capturar com suficiente precisão um grande volume de informações, bem como analisar a sua evolução temporal, e sua composição temática, sendo capaz de identificar sistematicamente as percepções de visitantes e moradores. O nosso estudo permitiu identificar na proteção e promoção do património cultural e natural, bem como das tradições das comunidades locais,

uma base sólida para o desenho de políticas públicas de turismo formuladas em termos de sustentabilidade. Por fim, a informação recolhida permitiu também uma análise qualitativa, permitindo identificar as principais ameaças e oportunidades à experiência do Caminho, bem como possíveis áreas de inovação em termos de produtos e serviços capazes de potenciar a experiência. O potencial desta última abordagem aconselharia a sua sistematização, talvez por meio da aplicação de técnicas da lexicometria (Iriarte, 2018, 2022) para fazer análises automatizadas do conjunto do corpus.

Autoria do trabalho

Conceptualização, (SSD, SSB); metodologia, (SSD, SBB); aquisição de dados, (SSD, SBB); análise e interpretação, (SSB, SSD); redação, revisão e edição (SSB, SSD). Todos os autores leram e concordam com a versão publicada do manuscrito.

Agradecimentos

Este trabalho foi financiado com cargo à *Convocatoria de 21 Proxectos de Investigación, Difusión e Didáctica sobre o Camiño de Santiago e as Peregrinacións*, promovido pela *Cátedra do Camiño de Santiago e das Peregrinacións* da Universidade de Santiago de Compostela.

VI. BIBLIOGRAFIA

Ali, Alisha e Frew, Andrew (2012). *Information and Communication Technologies for Sustainable Tourism*. Routledge. <https://doi.org/10.4324/9780203072592>

Ayeh, Julian K., Leung, Daniel, Au, Norman e Law, Rob (2012). Perceptions and Strategies of Hospitality and Tourism Practitioners on Social Media: An Exploratory Study. Em Matthias Fuchs, Francesco Ricci e Lorenzo Cantoni (eds.), *Information and Communication Technologies in Tourism 2012* (pp. 1–12). Springer Vienna.

Birkland, Thomas A. (2015). *An introduction to the policy process: Theories, concepts, and models of public policy making*. Routledge.

Dye, Thomas (1992). *Understanding Public Policy*, 7th ed. Prentice Hall.

Easton, David (1965). *A Framework for Political Analysis*. Prentice-Hall.

Fernández Méndez, Diana, Fernández Fernández, Melchor e Riveiro García, Dolores (2019). “Impacto socioeconómico del Camino de Santiago a nivel local: la percepción de los residentes”. *Documentos de Trabajo. Análise Económica*, 69, 1–39.

Gerston, Larry N. (2014). *Public policy making: Process and principles*. Routledge.

Gamallo, Pablo e García, Marcos (2017). *LinguaKit: Uma ferramenta multilingue para a análise linguística e a extração de informação*. *Linguamática*, 9(1), 19-28. <https://doi.org/10.21814/lm.9.1.243>

Torres Feijó, Elias J. (2020). *Bem-estar comunitário e visitantes através do Caminho em Santiago. Grandes narrativas, ideias e práticas culturais na cidade*. Andavira.

Flores-Ruiz, David, Elizondo-Salto, Adolfo, & de la O. Barroso-González, María (2021). Using Social Media in Tourist Sentiment Analysis: A Case Study of Andalusia during the Covid-19 Pandemic. *Sustainability*, 13, 3836.

Hall, Colin M., e Jenkins, John M. (1995). *Tourism and Public Policy*. Routledge.

Hennink, Monique e Kaiser, Bonnie N. (2019). Saturation in Qualitative Research. Em P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, e R.A. Williams (eds.), *SAGE Research Methods Foundations*. <https://dx.doi.org/10.4135/9781526421036822322>

Iriarte, Álvaro, Gamallo, Pablo e Simões, Alberto (2018). Estratégias Lexicométricas para Detetar Especificidades Textuais. *Linguamática*, 10, 19–26.

Iriarte, Álvaro e Sotelo, Susana (2022). Análise lexicométrica: Algumas técnicas aplicadas a entrevistas a visitantes de Santiago de Compostela. Em I. Caamaño, R. Samartim, & C. Lamela (eds.), *Cidades, turismo e caminhos de Santiago / Ciudades, turismo y caminos de Santiago*. Andavira.

McNamee, Paul (2005). Language Identification: A Solved Problem Suitable for Undergraduate Instruction. *Journal of Computing Sciences in Colleges*, 20(3), 94–101.

Medhat, Waala, Hassan, Ahmed, & Korashy, Hoda (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>

Mitjia T (2015). Evaluating language identification performance. *Engineering*. https://blog.twitter.com/engineering/en_us/a/2015/evaluating-language-identification-performance

Mittelstaedt, John D. e Mittelstaedt, Robert A. (2012). “Pilgrimage as Exchange: The Economy of Salvation and the Theory of Exchange”. *37th Macromarketing Conference*, 373–376.

Munar, Ana Maria e Jacobsen, Jens Kr. Steen (2014). Motivations for sharing tourism experiences through social media. *Tourism Management*, 43, 46–54. <https://doi.org/10.1016/j.tourman.2014.01.012>

OCS (2009). Informe do perfil do peregrino. Informe anual 2009. Informe Observatorio do Camiño de Santiago.

OCS (2017). Estudio do impacto socio-económico do Camiño de Santiago. Informe técnico. Observatorio do Camiño de Santiago.

Pearce, Douglas G. (1992). *Tourist Organisations*. Longman Scientific and Technical.

Putra, Rino S., Nurcahyo, Rahmat e Gabriel, Djoko S. (2019). Tourists Perception in Bali Using Social Media and Online Media Sentiment Analysis. *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1–5. <https://doi.org/10.1109/ICETAS48360.2019.9117317>

Sanderson, Ian (2002). Evaluation, policy learning and evidence-based policy making. *Public administration*, 80(1), 1-22.

Singh, Prabhsimran, Dwivedi, Yogesh K., Kahlon, Karanjeet S., Sawhney, Ravinder S., Alalwan, Ali A. e Rana, Nripendra P. (2020). Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing. *Information Systems Frontiers*, 22(2), 315–337. <https://doi.org/10.1007/s10796-019-09916-y>

Sotelo Docío, Susana e Benitez-Baleato, Jesus M. (2021). *Replication Data for: Digital Tracks: Application of Artificial Intelligence Technologies for Automatic detection of Perceptions from Social Media. The case of the Saint James Way, with a focus on COVID-19* (V1 ed.). Harvard Dataverse. <https://doi.org/10.7910/DVN/CUFZKT>

Xiang, Zheng, Du, Qianzhou, Ma, Yufeng e Fan, Weiguo (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>

Xunta de Galicia (2020). Ocupación hoteleira no Camiño Francés. Informe técnico. Área de Estudos e Investigación, Turismo de Galicia. <https://www.turismo.gal/aei/portal/index.php?idm=37>

Zeng, Benxiang e Gerritsen, Rolf (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27–36. <https://doi.org/10.1016/j.tmp.2014.01.001>

Zhang, Hailong, Gan, Wenyan e Jiang, Bo (2014). Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey. 2014 11th Web Information System and Application Conference, 262–265. <https://doi.org/10.1109/WISA.2014.55>

Zhang, Lei, Wang, Shuai e Liu, Bing (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>