Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail

Douglas Cirqueira[1][a], Fernando Almeida[2][b], Gültekin Cakir[4][c], Antonio Jacob[3][d], Fabio Lobato[2,3][e], Marija Bezbradica[1][f] and Markus Helfert[4][g]

[1] *School of Computing, Dublin City University, Dublin, Ireland*

[2] *Engineering and Geosciences Institute, Federal University of Western Pará, Santarém, Brazil*

[3] *Technological Sciences Center, State University of Maranhão, São Luís, Brazil*

[4] *Innovation Value Institute, Maynooth University, Maynooth, Ireland*

Keywords: Sentiment Analysis, Explainable Artificial Intelligence, Digital Retail, Crisis Management.

Abstract: Sentiment Analysis techniques enable the automatic extraction of sentiment in social media data, including popular platforms as Twitter. For retailers and marketing analysts, such methods can support the understanding of customers' attitudes towards brands, especially to handle crises that cause behavioural changes in customers, including the COVID-19 pandemic. However, with the increasing adoption of black-box machine learning-based techniques, transparency becomes a need for those stakeholders to understand why a given sentiment is predicted, which is rarely explored for retailers facing social media crises. This study develops an Explainable Sentiment Analysis (XSA) application for Twitter data, and proposes research propositions focused on evaluating such application in a hypothetical crisis management scenario. Particularly, we evaluate, through discussions and a simulated user experiment, the XSA support for understanding customer's needs, as well as if marketing analysts would trust such an application for their decision-making processes. Results illustrate the XSA application can be effective in providing the most important words addressing customers sentiment out of individual tweets, as well as the potential to foster analysts' confidence in such support.

1 INTRODUCTION

Crisis management and monitoring in social media are essential for retailers to understand their customers' needs (Mehta et al., 2020). A crisis in this context is defined as the negative reaction of customers towards particular products or services of a company, which can happen through their comments and messages on social media platforms (Vignal Lambret and Barki, 2018). That adverse reaction can impact organizations' reputation, as customers are increasingly adopting social media to reveal their opinions and sentiment on brands (Cirqueira et al., 2018).

Industry reports reveal the demanding profile of consumers on social media in 2019, for instance, with 78% of those who complain about a brand on Twitter expecting a response from the company within one hour[1]. Furthermore, the number of users consuming information over social media is increasing, which was also noticed during the current coronavirus crisis[2] (Sharma et al., 2020). Such a context highlights the need for retailers to interact with their customers and attend to their needs (de Almeida et al., 2017; Cirqueira et al., 2017). Furthermore, customer behavior has drastically changed with the corona pandemic, which can be classified as a moment of crisis (Donthu and Gustafsson, 2020).

Therefore, the question remains on how retailers can understand their customers' sentiment and needs in such scenarios. Sentiment Analysis (SA) methods, based on machine learning (ML) models, can sup-

[a] https://orcid.org/0000-0002-1283-0453
[b] https://orcid.org/0000-0003-4594-5924
[c] https://orcid.org/0000-0001-9715-7167
[d] https://orcid.org/0000-0002-9415-7265
[e] https://orcid.org/0000-0002-6282-0368
[f] https://orcid.org/0000-0001-9366-5113
[g] https://orcid.org/0000-0001-6546-6408

[1] https://www.lyfemarketing.com/blog/social-media-marketing-statistics/

[2] https://www.statista.com/statistics/1106766/media-consumption-growth-coronavirus-worldwide-by-country/

Cirqueira, D., Almeida, F., Cakir, G., Jacob, A., Lobato, F., Bezbradica, M. and Helfert, M.
Explainable Sentiment Analysis Application for Social Media Crisis Management in Retail.
DOI: 10.5220/0010215303190328
In *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2020)*, pages 319-328
ISBN: 978-989-758-480-0

port such understanding (Hasan et al., 2018). However, those methods are usually black-boxes (Adadi and Berrada, 2018), and provide only a score or label for a document being classified as positive or negative. Retailers would need more and understand what factors impact customers' opinion and their attitudes towards their brands (Micu et al., 2017).

Explainable AI (XAI) aims to develop explanation methods, which enable transparency and understanding of ML models while keeping their good learning performance (Miller, 2019; Cirqueira et al., 2020). Indeed, researchers have considered the intersection between SA and XAI research, generating Explainable Sentiment Analysis applications (XSA), supporting stakeholders in understanding the sentiment and reasoning behind predictions for enhancing the interaction between decision-makers and SA applications (Mathews, 2019).

Among social media platforms, Twitter is a valuable example and recognized for revealing the real state of word of mouth globally (Jansen et al., 2009; Vo et al., 2019). However, there is a lack of investigation on the adoption of XSA for supporting retailers dealing with crisis management on Twitter. Research shows the value of that platform as a data source for social media analytics and insights, which could also support the handling of crises and to react to consumers' needs (Stieglitz et al., 2018). Nevertheless, the assessment of XSA in this context is essential as retailers would need to trust and rely on such technology for their decision-making and strategies.

Therefore, this study introduces the following research propositions (RP):

1. RP1: An XSA application can support retailers and analysts in managing social media crises by understanding customer sentiment and needs on Twitter
2. RP2: Retailers and analysts would trust an XSA application for decision-making, which can be evaluated based on XAI evaluation frameworks

This study presents prior research to support the given propositions, and develops an XSA application to be evaluated in the context of crisis management on Twitter. The study illustrates a hypothetical crisis management scenario for a retailer dealing with customers' negative attitudes via Twitter. The two research propositions are reflected based on discussions and a simulated user experiment, leveraging the XSA predictions for customers' sentiment on a Twitter dataset.

The rest of this paper is organized as follows: Section 2 provides related work within the core themes of this study; Section 3 describes the crisis management scenario adopted in the study; Section 4 details the development of the Explainable Sentiment Analysis application; Section 5 discusses the research propositions and evaluation of the XSA application; Section 6 presents final remarks and future work directions.

2 RELATED WORK

2.1 Crisis Management

Crisis management is a critical process, and companies need to undergo in cases of unprecedented and disruptive events, potentially harming the existence of whole organisations (Bundy et al., 2017). Organisational competence requires the ability to detect signals early in order to sense potential threats and act accordingly (James and WOOTEN, 2005). For instance, the COVID-19 pandemic represents a typical crisis event, especially for the retailing sector (Pantano et al., 2020).

Social media can play a crucial role in detecting disruptive threats and understanding changing consumer behaviour (Saroj and Pal, 2020; Mehta et al., 2020). Therefore, researchers have highlighted social media platforms as a fertile ground for crises. Some have performed empirical investigations and manual analysis of customers' emotions online to obtain insights for mitigating threats to their brands (Vignal Lambret and Barki, 2018). Others have focused on the challenges and opportunities enabled by those platforms, and emphasized the need for solutions that help in sensing and decision-making and usability to support crisis managers (Zhu et al., 2017; Stieglitz et al., 2018).

2.2 Twitter as a Source of Opinions

Researchers recognize twitter as a valuable source of the current state of word of mouth globally (Vo et al., 2019). The platform is known for being dynamic and for the speed in which tweets are generated, spreading the message and attitudes of users towards a wide variety of topics (Steinskog et al., 2017). For instance, the latest statistics on Twitter usage point out to 6000 tweets being sent per second[3]. Political parties, financial institutions, and policymakers have already explored this platform's potential to understand the public and forecast the behavior of a population or the market (Bovet et al., 2018; Pagolu et al., 2016).

Companies and marketing teams also realized the need for a presence on such a platform, where users

[3]https://www.dsayce.com/social-media/tweets-day/

present their opinions on other users and brands and their services (Rathan et al., 2018). Researchers in Social Customer Relationship Management (Social CRM) also discuss this network's potential for integrating customer engagement on social media within traditional CRM platforms (Lobato et al., 2016). The platform features also facilitate the investigation through its data, based on Application Programming Interfaces (API) and friendly access to developers (Liu, 2019).

2.3 Sentiment Analysis

Sentiment Analysis methods and techniques aim to automatically detect the sentiment, opinion, and polarity present in textual datasets (Liu, 2012; Cirqueira et al., 2016). SA methodologies are diverse, and include, for instance, dictionary-based and ML-based methods, the last being widely explored and recognized as efficient in the field (Medhat et al., 2014). SA supports a wide range of applications, including the retail sector, especially in the domain of social media monitoring, where user-generated content can reveal trends and customers' attitudes towards brands and companies (Hu et al., 2017).

2.4 Explainable AI

Explainable AI (XAI) researchers aim to keep the learning capacity of ML models, while enabling their transparency and understanding for its users (Miller, 2019; Holzinger et al., 2020). Such understanding is enabled through explanations, which can be intrinsic to transparent ML models, such as decision trees, or provided by model-agnostic and post-hoc methods (Molnar, 2019).

Within the disciplines of computer science and ML, the development of model-agnostic explanations has been popular. Examples include LIME (Ribeiro et al., 2016) and SHAP (Lundberg and Lee, 2017). Both methods enable the visualization of feature importance and impact on predictions of AI models, which has proven to be useful for decision-making in a diversity of scenarios (Slack et al., 2019). Such explanations can be provided at a local level, when explaining a particular prediction, or global level, when explaining the whole logic of an AI model (Adadi and Berrada, 2018).

2.5 Explainable Sentiment Analysis

These studies aim not only the extraction and predictions for the sentiment, but also explanations for why a particular polarity is attributed to a textual dataset instance. Such applications have supported researchers and decision-makers to consider a human-in-the-loop perspective to enhance SA techniques, and evaluate existing explanation methods through SA applications (Zucco et al., 2018; So, 2020; Hase and Bansal, 2020; Silveira et al., 2019).

Regarding Twitter data, researchers have investigated XSA support in understanding the behavior of voters (Mathews, 2019). Others have explored the adoption of explanation methods for mining tweets topics and automatic text generation (Islam, 2019; Ehsan et al.,). However, the case of XSA for supporting retailers in understanding customers on Twitter is rarely observed, especially in crisis management contexts.

3 CRISIS MANAGEMENT SCENARIO

We develop a hypothetical scenario, inspired in previous references and case studies (Saroj and Pal, 2020; Mehta et al., 2020; Vignal Lambret and Barki, 2018; Zhu et al., 2017; Stieglitz et al., 2018). The context represents a crisis management scenario, where a retailer or marketing analyst needs to undergo through tweets discussing issues on their products. In the case of the adopted dataset for this study, tweets are about electronic products, including laptops and smartphones of a big company.

Companies usually have a department focused on monitoring the reputation of the brand. That department is composed of professionals aiming to provide retailers with the feedback and needs of brand consumers (Tsirakis et al., 2017). In the hypothetical scenario, there is the release of a new series of electronic products, and customers have been discussing the new features. A Twitter dataset representing such a scenario has been identified, and employed in the experiments within this study[4].

However, as expected, not all the comments and tweets are positive, and the company needs to react to the negative observations of customers, and identify their reasons. Consumers' negative comments are known for influencing customers' attitudes searching for information on a particular brand or product (Baek et al., 2014). To keep a good image and reputation is vital for the company in this scenario as well.

Therefore, in this scenario, marketing analysts would have two tasks: to analyze customers' sentiment on those tweets and understand their particular needs. In the end, they would report their findings

[4]https://data.world/crowdflower/apple-twitter-sentiment

back to the responsible department with their company. Such an analysis could hardly be performed manually, given the large number of tweets available. Therefore, those professionals would have the support of an XSA application.

The XSA application automatically provides the scores for sentiments that can be detected on customers' tweets, as well as explanations for the predictions. However, for relying on such predictions, those professionals need to rely on and trust them (Zhang et al., 2020). Therefore, the next sections present the development and evaluation of an XSA application, regarding the two tasks of those professionals in this scenario, which are aligned with our research propositions.

4 EXPLAINABLE SENTIMENT ANALYSIS APPLICATION

The development stages of this research follow a CRISP-DM methodology, composed of the following steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Wirth and Hipp, 2000). For business and data understanding, it is reviewed the literature in XSA and explored available datasets within the Twitter domain, which could represent a crisis management scenario for a particular retailer.

Therefore, a dataset of tweets discussing electronics products is identified, which is considered representative of our context and mentioned in Section 3. It is adopted positive and negative instances out of this dataset, resulting in 375 positive and 1092 negative tweets, a total of 1467 instances. For data preparation, it is regarded literature of preprocessing steps for SA applications (Cirqueira et al., 2018; Krouska et al., 2016). The implemented preprocessing steps are lowercasing, stopwords removal, stemming, special character removal, punctuation removal, numbers removal, and emojis removal, widely explored in SA literature. The character "#" for hashtags is removed, and the hashtag content is kept. Mentions and Twitter handles are also removed.

For modeling, we use the mentioned dataset to train and test the ML models of Support Vector Machines (SVM) (Steinwart and Christmann, 2008), Random Forest (RF) (Breiman, 2001), XGB Boost Classifier (Chen and Guestrin, 2016), and Neural Network Multilayer Perceptron (MLP) (Popescu et al., 2009). Those are well regarded as good performing in the field of SA (Luo et al., 2016; Al Amrani et al., 2018; Jabreel and Moreno, 2018). We perform cross-validation with 5 folds (Browne, 2000), and observe the MLP model performs the best regarding the F1 score (Zhang et al., 2015). Table 1 illustrates the MLP performance compared to the other tested models. Our MLP model contains one hidden layer with 150 neurons, and an output layer with two neurons for the positive and negative sentiment classes. Therefore, for the next development steps, the MLP model is adopted.

Table 1: Average F1 Scores for 5-Fold Cross Validation on the Twitter Dataset.

ML Model	Average F1 Score
SVM	0.75
RF	0.74
XGBoost	0.79
MLP	0.81

Furthermore, to provide explanations within our XSA application, LIME and SHAP's explanation methods are employed, given their well-documented material and wide adoption in research and industry. Therefore, this study assumes such methods reflect the current state-of-art and practice in XAI research, which guarantees our application is up-to-date in this regard. Therefore, based on results in Table 1, we selected the MLP model to be trained and explained based on the Twitter dataset and selected explanation methods. For training the chosen MLP model and providing explanations, we preprocess and split the data in train and test (85/15 ratio). It is assessed again the F1 score obtained, which is 89%.

Figure 1 depicts the main components of the application developed, from the training and testing of AI models to the explanation of their predictions.

5 DISCUSSION AND EXPERIMENTS

5.1 RP1: An XSA Application Can Support Retailers and Analysts in Managing Social Media Crises by Understanding Customer Sentiment and Needs on Twitter

Figures 2 depicts the explanations provided by the selected explanation methods for the sentiment of customers on Twitter, and that an analyst can visualize out of our application. The support for research proposition one is discussed based on how retailers could leverage such explanations to understand their customer needs during a crisis. First of all, an

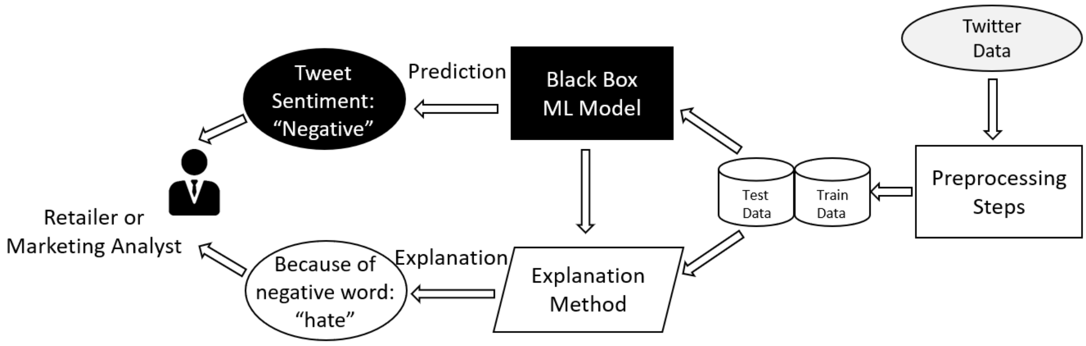

Figure 1: Explainable sentiment analysis application.

XSA application will reduce retailers' or marketing analysts' burden to manually assess users' and customers' opinions online. Furthermore, the application automatically provides the scores for sentiments that can be detected on customers' tweets.

In the example of Figure 2, a customer complains about a particular feature of his recently bought smartphone, following the pattern described in the scenario presented in Section 3. It is fundamental for the manufacturer to identify the reason for this complaint, and react whether through its internal processes or provide the customer with some support or feedback. The complaint might have the power to influence other customers' willingness to buy the same product. In this example, the customer complains that his device does not include the latest software update, although it is a brand new product.

Regarding the support of explanation methods, Figure 2.A illustrates explanations of local feature importance (LFI XSA), enabled by the LIME tool. With these explanations, a retailer would be able to analyse which words within a tweet are the most important for predicting sentiment. With this explanation, retailers might increase their confidence in an AI partner for SA, as they would be able to review if the words selected as important by an AI model for positive or negative sentiments are in line with their own experience concerning customers' feedback. In this case, the explanation method detects as negative words "without", "stupid", "stuff", "software", "sent", "iphon", and "latest". From these words, it can be noticed that the method says the most important negative word is "without", which can already give a clue to the marketing analyst on what type of problem a customer refers to, and the lack of a software update in this case.

In Figure 2.B, a retailer would perceive the most important words and keywords being used to provide them with predictions (GFI XSA). Those are enabled by the global feature importance method, which is instantiated through the SHAP tool. With this explanation, retailers have in their hands the words they should pay attention to whenever analysing the comments and tweets of their customers online. They can think of particular strategies and prioritize customers' support, based on the usage of those keywords in their feedback and interaction with the company.

Finally, in Figure 2.C, retailers can perceive the impact of words pushing predictions for positive and negative sentiments (FI XSA). It also shows if words not present in the tweet can impact the prediction. In our case, the method reinforces some of the important words for a negative prediction as "without", "stupid", "stuff", "software", which echoes the method in Figure 2.A. This explanation can support those retailers in confirming if the words selected by AI models for particular sentiments match their experience and the impact those have on such predictions. Retailers can consider such output for detecting the particular needs of their customers. In this context, it would be highlighted for a marketing analyst the need to look into issues with the customer smartphone software. Furthermore, customers can complain about particular product features or aspects of a previously received service.

5.2 RP2: Retailers and Analysts Would Trust an XSA Application for Decision-making, Which Can Be Evaluated based on XAI Evaluation Frameworks

Besides discussing the potential of XSA in supporting marketing analysts, it would be important for those experts to trust such support. To evaluate our XSA concerning that aspect, we discuss the evaluation of explanations and explanation methods in XAI, which is still an ongoing researched topic. However, the work of (Doshi-Velez and Kim, 2017) provides

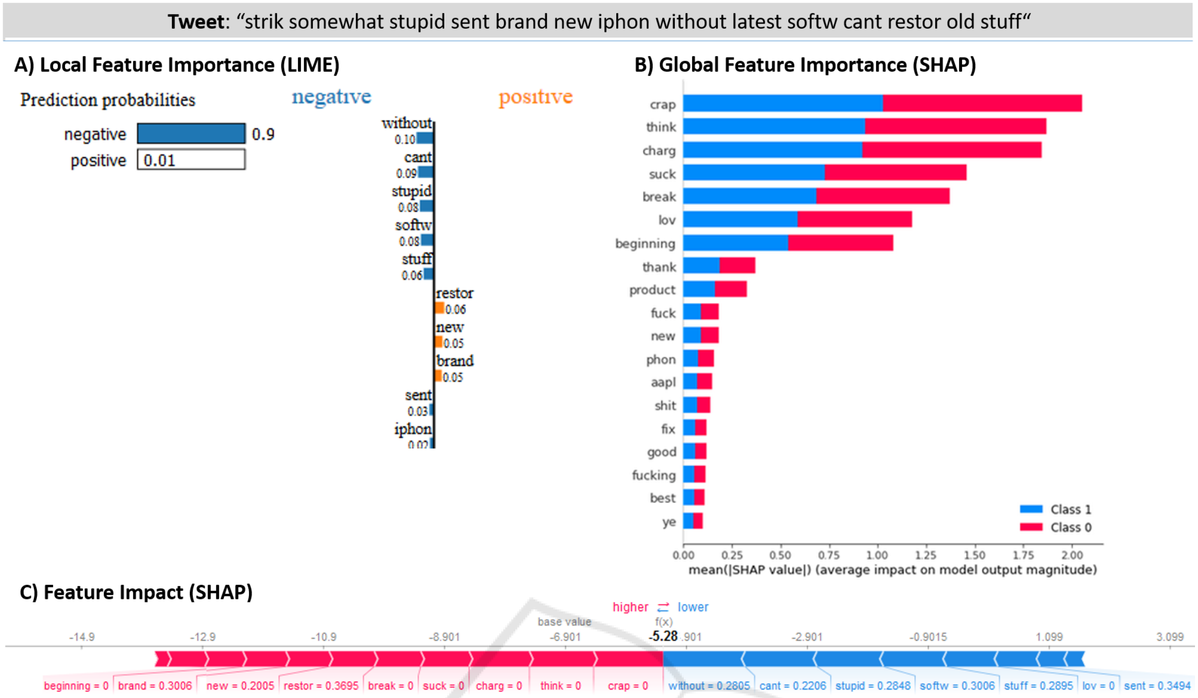

Figure 2: Explanation methods for analysis of sentiment on Twitter.

a robust methodology for the evaluation of explanations, based on three strategies. First, a functionally-grounded evaluation of the interface, in which a researcher defines proxy tasks for assessing how good an explanation is in achieving its goal, without human participation (Sokol and Flach, 2020). In the context of XSA in retail, it would be essential to assess retailers' level of trust or confidence towards explanations. Therefore, a simulated user experiment to evaluate the level of trust can be performed. This approach has been previously adopted and accepted for evaluating explanation methods in XAI literature (Ribeiro et al., 2016; Weerts et al., 2019; Nguyen, 2018; Honegger, 2018).

Second, a human-grounded evaluation of XSA. In this case, real human participants would be required, but simple tasks could be established to be performed by the participants. That opens opportunities for the recruitment of users that are not real retailers, who could be immersed in a social media monitoring scenario for retail. The last evaluation approach by (Doshi-Velez and Kim, 2017) is named application-grounded. In this case, real human participants should go through real tasks for a particular domain. In the case of XSA for retailers, it would be required to account for their real tasks when monitoring social media customers.

For this study, we adopt a functionally-grounded evaluation strategy. That enables the development of our simulated user experiment. The aim is to estimate the confidence of users on the explanation methods provided by our XSA application. Therefore, following a validated simulation approach (Nguyen, 2018), for the explanation methods in Figure 2.A and 2.C, we estimate the user confidence based on the average switching point (ASP) for a prediction, when deleting words in their order of importance. The importance is given by the explanation methods. Thus, the lower the ASP, the better, as it proves the methods were able to detect the most important words to an expert assess tweets sentiment and needs. The results compare the deletion by order of importance to a random selection of words to be deleted. Table 2 shows the average results of this experiment, over ten runs with different random seeds for deleting words (LFI XSA Random and FI XSA Random).

We have different scenarios based on different levels of confidence of the ML model on predictions. We wanted to evaluate how the ML confidence might influence on the ASP, and user confidence. We also filtered out tweets with less than three words, given our analysis is focused on their removal, and with fewer words, results could not reflect the benefit of XSA feature importance against a random selection.

Table 2 shows that the ASP is always lower when deleting words based on their order of importance

Table 2: Results for Simulated User Experiment and Average Switching Point when Deleting Words and Checking Predictions (0.0 - 0.14 = 1 word / 0.15 - 0.24 = 2 words / 0.25 - 0.34 = 3 words / 0.35 - 0.40 = 4 words).

ML Model Confidence	Average Switching Point			
	LFI XSA	LFI XSA Random	FI XSA	FI XSA Random
95%	0.16	0.33	0.14	0.40
90%	0.15	0.30	0.15	0.38
85%	0.13	0.22	0.13	0.37
80%	0.12	0.30	0.09	0.25

provided by explanation methods compared to a random deletion. Notably, an average maximum of two words would need to be deleted for changing the prediction when doing so. When deleting words randomly, we need to delete a minimum of 3 to 4 words to change the prediction, except for the scenario with 85% of ML confidence, where the LFI XSA Random scores 0.22 ASP. That shows retailers and marketing analysts would be provided with the most important words by explanation methods on average, reflecting the importance of which aspects they need to consider for identifying consumer sentiment and needs.

Second, for evaluating the explanation method in Figure 2.B, we follow a similar approach as adopted for 2.A and 2.C. However, given the important words by this method are valid for the full set of predictions, we delete the words by their order of importance from all the test data instances, and check the impact on the F1 score for the complete test dataset. However, Figure 3 shows no significant difference when randomly deleting words compared to deleting words by their order of importance, providing F1 scores of 0.84 and 0.88, respectively. Therefore, for this method, further experiments and a bigger dataset would be suggested for evaluating its feasibility in estimating user trust within a crisis management scenario. We also believe that a fine-grained approach could be undertaken for evaluating this explanation method. For instance, tweets could be clustered based on their particular topics and keywords, for which this evaluation could be performed separately, according to the top important and impactful keywords identified by explanation methods.

Therefore, we argue that a retailer and marketing analysts might trust, from an estimated confidence perspective, the predictions and explanations out of the XSA application, given the low ASP provided when relying on the importance and impact of words. However, further experiments are needed, which can have as impacting factors the size and types of explanation methods adopted.

6 FINAL REMARKS

This study develops an Explainable Sentiment Analysis application and provides two research propositions regarding the adoption of the application in the context of Twitter and crisis management for retailers, and assesses the support for those propositions through discussions and simulated user experiments. We presented an application that could be adopted to provide retailers with explanations and insights on their user needs during social media crises.

For researchers and practitioners interested in the field, we discussed a use case demonstrating how this relationship between XSA and retailers could be depicted, and the potential of XSA support. Furthermore, the simulated user experiments highlight the potential impact on user confidence for working with an XSA when assessing user feedback on social media. Moreover, to consider a user in the loop perspective, and feeding the XSA system back with misclassifications, could help in investigating the impact of a human-centered perspective for XSA in crisis management.

Although we discuss and conduct simulations, the adoption of user experiments based on the two strategies discussed by (Doshi-Velez and Kim, 2017) of human and application-grounded approaches, could be useful to assess the impact of an XSA on the daily decision-making of marketing analysts. The elicitation of a crisis management scenario with marketing analysts would also enrich future studies in the area.

As future work, from the perspective of experiments, we point out simulations with bigger and multiple datasets in crisis management. Experiments could also be performed with Recurrent Neural Networks and word embeddings, enhancing ML predictions' performance, and potentially the estimated confidence on explanation methods results.

Furthermore, the research propositions could be turned out into hypotheses or tested through user studies for their assessment. A user-centric and information systems perspective could also be considered for developing an XSA, by first performing requirements

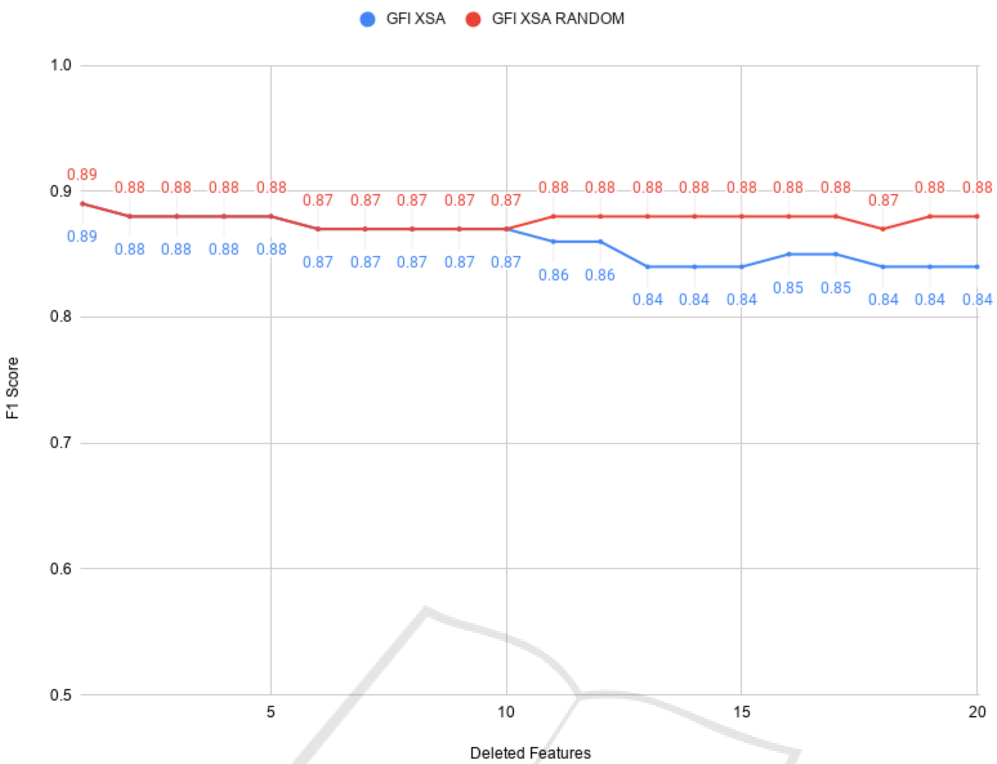

Figure 3: Results for Feature Deletion by Global Feature Importance and Random Selection.

elicitation of users for deploying explanation methods attending to their needs and tasks for SA. With such elicitation, requirements might emerge to different methods, such as case-based explanations. A literature review on the capabilities of explanation methods to SA would also be a further avenue within those disciplines' intersection. Finally, a research agenda could be derived from research gaps and opportunities in the intersection of SA and XAI research, emphasizing the decision-making support to end-users, domain experts, and developers.

ACKNOWLEDGEMENTS

This research was supported by the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 765395; and supported, in part, by Science Foundation Ireland grant 13/RC/2094.

REFERENCES

Adadi, A. and Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160.

Al Amrani, Y., Lazaar, M., and El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127:511–520.

Baek, H., Ahn, J., and Oh, S. (2014). Impact of tweets on box office revenue: focusing on when tweets are written. *ETRI Journal*, 36(4):581–590.

Bovet, A., Morone, F., and Makse, H. A. (2018). Validation of twitter opinion trends with national polling aggregates: Hillary clinton vs donald trump. *Scientific reports*, 8(1):1–16.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Browne, M. W. (2000). Cross-validation methods. *Journal of mathematical psychology*, 44(1):108–132.

Bundy, J., Pfarrer, M. D., Short, C. E., and Coombs, W. T. (2017). Crises and crisis management: Integration, interpretation, and research development. *Journal of Management*, 43(6):1661–1692.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Cirqueira, D., Jacob, A., Lobato, F., de Santana, A. L., and Pinheiro, M. (2016). Performance evaluation of sentiment analysis methods for brazilian portuguese. In *International Conference on Business Information Systems*, pages 245–251. Springer.

Cirqueira, D., Nedbal, D., Helfert, M., and Bezbradica, M. (2020). Scenario-based requirements elicitation for user-centric explainable ai. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pages 321–341. Springer.

Cirqueira, D., Pinheiro, M., Braga, T., Jacob Jr, A., Reinhold, O., Alt, R., and Santana, Á. (2017). Improving relationship management in universities with sentiment analysis and topic modeling of social media channels: learnings from ufpa. In *Proceedings of the International Conference on Web Intelligence*, pages 998–1005.

Cirqueira, D., Pinheiro, M. F., Jacob, A., Lobato, F., and Santana, A. (2018). A literature review in preprocessing for sentiment analysis for brazilian portuguese social media. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 746–749. IEEE.

de Almeida, G. R., Cirqueira, D. R., and Lobato, F. M. (2017). Improving social crm through eletronic word-of-mouth: a case study of reclameaqui. In *Anais Estendidos do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 107–110. SBC.

Donthu, N. and Gustafsson, A. (2020). Effects of covid-19 on business and research. *Journal of business research*, 117:284.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Ehsan, U., Purdy, C., Kelley, C., Polepeddi, L., and Davis, N. Unpack that tweet: A traceable and interpretable cognitive modeling system.

Hasan, A., Moin, S., Karim, A., and Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, 23(1):11.

Hase, P. and Bansal, M. (2020). Evaluating explainable ai: Which algorithmic explanations help users predict model behavior? *arXiv preprint arXiv:2005.01831*.

Holzinger, A., Carrington, A., and Müller, H. (2020). Measuring the quality of explanations: the system causability scale (scs). *KI-Künstliche Intelligenz*, pages 1–6.

Honegger, M. (2018). Shedding light on black box machine learning algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions. *arXiv preprint arXiv:1808.05054*.

Hu, G., Bhargava, P., Fuhrmann, S., Ellinger, S., and Spasojevic, N. (2017). Analyzing users' sentiment towards popular consumer industries and brands on twitter. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 381–388. IEEE.

Islam, T. (2019). Ex-twit: Explainable twitter mining on health data. *arXiv preprint arXiv:1906.02132*.

Jabreel, M. and Moreno, A. (2018). Eitaka at semeval-2018 task 1: An ensemble of n-channels convnet and xgboost regressors for emotion analysis of tweets. *arXiv preprint arXiv:1802.09233*.

James, E. and WOOTEN, L. (2005). Leadership as (un)usual:: How to display competence in times of crisis. *Organizational Dynamics*, 34:141–152.

Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188.

Krouska, A., Troussas, C., and Virvou, M. (2016). The effect of preprocessing techniques on twitter sentiment analysis. In *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–5. IEEE.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Liu, X. (2019). A big data approach to examining social bots on twitter. *Journal of Services Marketing*.

Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2016). Social crm: Biggest challenges to make it work in the real world. In *International Conference on Business Information Systems*, pages 221–232. Springer.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774.

Luo, F., Li, C., and Cao, Z. (2016). Affective-feature-based sentiment analysis using svm classifier. In *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 276–281. IEEE.

Mathews, S. M. (2019). Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review. In *Intelligent Computing-Proceedings of the Computing Conference*, pages 1269–1292. Springer.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.

Mehta, S., Saxena, T., and Purohit, N. (2020). The new consumer behaviour paradigm amid covid-19: Permanent or transient? *Journal of Health Management*, 22(2):291–301.

Micu, A., Micu, A. E., Geru, M., and Lixandroiu, R. C. (2017). Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology & Marketing*, 34(12):1094–1100.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

Molnar, C. (2019). Interpretable machine learning. *Lulu. com*.

Nguyen, D. (2018). Comparing automatic and human evaluation of local explanations for text classification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1069–1078.

Pagolu, V. S., Reddy, K. N., Panda, G., and Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE.

Pantano, E., Pizzi, G., Scarpi, D., and Dennis, C. (2020). Competing during a pandemic? retailers' ups and downs during the covid-19 outbreak. *Journal of Business Research*.

Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., and Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588.

Rathan, M., Hulipalled, V. R., Venugopal, K., and Patnaik, L. (2018). Consumer insight mining: Aspect based twitter opinion mining of mobile phone reviews. *Applied Soft Computing*, 68:765–773.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144.

Saroj, A. and Pal, S. (2020). Use of social media in crisis management: A survey. *International Journal of Disaster Risk Reduction*, 48:101584.

Sharma, K., Seo, S., Meng, C., Rambhatla, S., and Liu, Y. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.

Silveira, T. D. S., Uszkoreit, H., and Ai, R. (2019). Using aspect-based analysis for explainable sentiment predictions. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 617–627. Springer.

Slack, D., Hilgard, S., Jia, E., Singh, S., and Lakkaraju, H. (2019). How can we fool lime and shap? adversarial attacks on post hoc explanation methods. *arXiv preprint arXiv:1911.02508*.

So, C. (2020). What emotions make one or five stars? understanding ratings of online product reviews by sentiment analysis and xai. *arXiv preprint arXiv:2003.00201*.

Sokol, K. and Flach, P. (2020). Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 56–67.

Steinskog, A., Therkelsen, J., and Gambäck, B. (2017). Twitter topic modeling by tweet aggregation. In *Proceedings of the 21st nordic conference on computational linguistics*, pages 77–86.

Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Stieglitz, S., Mirbabaie, M., Fromm, J., and Melzer, S. (2018). The adoption of social media analytics for crisis management–challenges and opportunities.

Tsirakis, N., Poulopoulos, V., Tsantilas, P., and Varlamis, I. (2017). Large scale opinion mining for social, news and blog data. *Journal of Systems and Software*, 127:237–248.

Vignal Lambret, C. and Barki, E. (2018). Social media crisis management: Aligning corporate response strategies with stakeholders' emotions online. *Journal of Contingencies and Crisis Management*, 26(2):295–305.

Vo, T. T., Xiao, X., and Ho, S. Y. (2019). How does corporate social responsibility engagement influence word of mouth on twitter? evidence from the airline industry. *Journal of Business Ethics*, 157(2):525–542.

Weerts, H. J., van Ipenburg, W., and Pechenizkiy, M. (2019). Case-based reasoning for assisting domain experts in processing fraud alerts of black-box machine learning models. *arXiv preprint arXiv:1907.03334*.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Springer-Verlag London, UK.

Zhang, D., Wang, J., and Zhao, X. (2015). Estimating the uncertainty of average f1 scores. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 317–320.

Zhang, Y., Liao, Q. V., and Bellamy, R. K. (2020). Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 295–305.

Zhu, L., Anagondahalli, D., and Zhang, A. (2017). Social media and culture in crisis communication: Mcdonald's and kfc crises management in china. *Public Relations Review*, 43(3):487–492.

Zucco, C., Liang, H., Di Fatta, G., and Cannataro, M. (2018). Explainable sentiment analysis with applications in medicine. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1740–1747. IEEE.