A Pipeline for Measuring Brand Loyalty Through Social Media Mining

A Pipeline for Measuring Brand Loyalty Through Social Media Mining

Hazem Samoaa[1(✉)] and Barbara Catania[2]

[1] Chalmers University of Technology, Gothenburg, Sweden
samoaa@chalmers.se
[2] University of Genoa, Genoa, Italy
barbara.catania@unige.it

Abstract. Enhancing customer relationships through social media is an area of high relevance for companies. To this aim, Social Business Intelligence (SBI) plays a crucial role by supporting companies in combining corporate data with user-generated content, usually available as textual clips on social media. Unfortunately, SBI research is often constrained by the lack of publicly-available, real-world data for experimental activities. In this paper, we describe our experience in extracting social data and processing them through an enrichment pipeline for brand analysis. As a first step, we collect texts from social media and we annotate them based on predefined metrics for brand analysis, using features such as sentiment and geolocation. Annotations rely on various learning and natural language processing approaches, including deep learning and geographical ontologies. Structured data obtained from the annotation process are then stored in a distributed data warehouse for further analysis. Preliminary results, obtained from the analysis of three well known ICT brands, using data gathered from Twitter, news portals, and Amazon product reviews, show that different evaluation metrics can lead to different outcomes, indicating that no single metric is dominant for all brand analysis use cases.

1 Introduction

Improving customer relationships through social media, and using information that can be learned from social media content to further improve products, is an area of high relevance for companies. Traditionally, companies rely on their internal Customer Relationship Management (CRM) systems to measure customer satisfaction. However, existing CRM approaches are often not sufficient to analyze and understand customer behaviour. Frequently, feedback from external systems, such as social media, is crucial to truly understand a brand's image. To this end, Social Business Intelligence (SBI) combines corporate data with user-generated content (UGC) to make decision-makers aware of important brand-related trends, and to improve decision making through timely feedback [2,6]. UGC attempts to capture brand image through mining online textual clips, comprising the tastes, opinions, feedback, and actions from customers and other

T. Bureš et al. (Eds.): SOFSEM 2021, LNCS 12607, pp. 489–504, 2021.
https://doi.org/10.1007/978-3-030-67731-2_36

stakeholders alike. Textual clips span from messages posted on social media or articles taken from online newspapers or magazines to customer reviews collected from reviews portal such as Amazon or the Google Appstore.

Extracting useful information from textual UGC requires first crawling data sources to acquire relevant clips, followed by enrichment. Enrichment activities may entail simply identifying structured metadata (e.g., date, community, or review score), or using natural language processing (NLP) techniques to find the relevant concepts the clip mentions (e.g., the brands) and, if possible, to assign a sentiment (i.e., positive, negative, or neutral) to it [16]. SBI thus poses research challenges in many areas, including information retrieval, data mining, and NLP. Unfortunately, SBI research is often constrained by the lack of publicly-available, real-world data for experimental activities [2]. This might limit research results in this field.

In this paper, we describe our experience in extracting social data and processing them through an enrichment pipeline that produces valuable annotations for *brand loyalty analysis*.

As a first contribution, we present the system we have designed for collecting social media data and storing them in a data warehouse for further brand loyalty analysis with respect to many different dimensions. Texts are collected from different social media (Twitter, news portals, and Amazon product reviews) and annotated with features such as sentiment and geolocation. The obtained structured data are then stored, together with stock price information, in a distributed data warehouse based on the Hadoop File System (HDFS) and Spark SQL. The data warehouse has been designed for supporting brand loyalty analysis with respect to many different dimensions. The measures correspond to key performance indicators (KPIs) related to customer experience, customer satisfaction, and customer interaction.

As a second contribution, to show the flexibility of the proposed approach, we present some of the analytical results obtained from the analysis of three well known ICT brands, discussing the impact of the selected data sources on the whole process. The obtained results show that different evaluation metrics can lead to different outcomes, indicating that no single metric is dominant for all brand analysis use cases.

The remainder of this paper is organized as follows. In Sect. 2 we briefly review related work. The pipeline we propose is presented in Sect. 3 while details about the data annotation processes are described in Sect. 4. Section 5 presents the data warehouse design. Some preliminary experimental results are then discussed in Sect. 6, to show the applicability of the proposed pipeline. Finally, Sect. 7 presents some conclusions and outlines future work.

2 Related Work

SBI has been investigated in many different domains and many proposals exists to cope with social data and UGC in a business intelligence context. To just name a few, [14] proposes an SBI approach for analyzing term occurrences in

documents belonging to a corpus. [22] presents textual measures as a solution to summarize textual information. In [24], a complete architecture for OLAP analysis of tweets in order to gain statistical information has been presented. In this paper, we build on these existing SBI proposals to develop a pipeline for an SBI approach to brand analysis.

UGC often comes in the form of textual clips. Tweets, news articles, and reviews are just some examples of types of clips that could be of interest for SBI platforms. A core issue is how to extract structured information, such as customer sentiment and geolocation, from textual clips. For *customer sentiment*, many NLP approaches have been proposed, e.g., [19]. Other approaches rely on deep learning techniques. One of the first works in this direction investigates the effects of quality, value, and customer satisfaction on consumer behavioral intention through various machine learning and deep learning classification algorithms [3]. A more recent work proposes a generic way to understand the behavioral intention of the customer [27]. These works tackle customer content for quite general purposes ("voice of customer") without aiming for metrics to extract measures and compare brands. On the other hand, we rely on a combination of different approaches for understanding customer opinions and measuring, through specific metrics, the strength of the relationships between the brand and the consumers.

A common intuition is that users often mention places that are near their current location. Several approaches have been presented to automatically *geolocate* non-geotagged textual clips using textual content [5,7,11,23,29]. Most of these methods rely on a training phase, during which they construct language models, in order to probabilistically infer the location of unseen messages. These types of models can very accurately geolocate microblog messages at a city level [5,11], but suffer from problems related to text noise, such as misspellings, non-textual content, or the presence of links. Moreover, during classification, the finer-grained the grid used to geolocate is (i.e., the higher the sub-city detail), the higher the number of classes which negatively affects performance. To overcome these problems, recent alternative approaches have been proposed that exploit external information sources publicly available on the web, e.g., GeoNames or OpenStreetMaps, to reach sub-city accuracy [25]. In our work, we exploit both data and knowledge-driven approaches for geolocating textual clips.

3 A Pipeline for Brand Analysis

A high-level schematic overview of the proposed pipeline is given in Fig. 1. As a first step, we crawl three different types of data sources to retrieve different types of textual clips, namely tweets from Twitter, news, and reviews from Amazon. Collected textual clips are then annotated by extracting both implicit and explicit structured information. Explicit features refer to the reference *time* and the *brand* of the clip. Time is usually explicitly associated with textual clips while brand information can be extracted through simple keyword search. Implicit features correspond to *sentiment* information (which is fundamental when mining customer opinion), and *geolocation*, when it is not explicitly provided. The

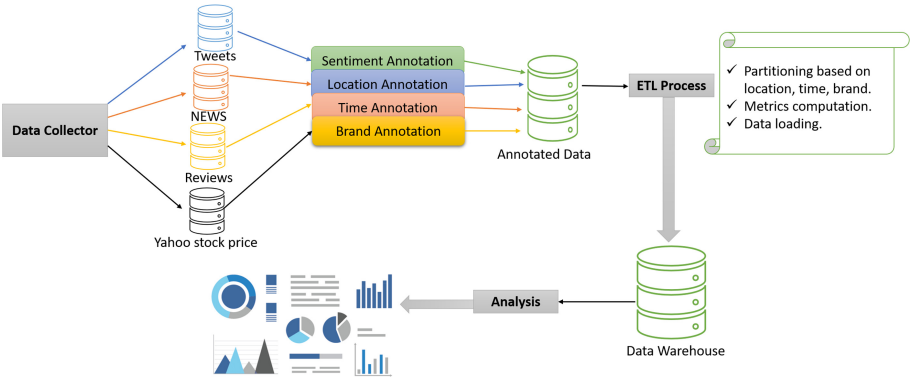

Fig. 1. SBI Pipeline for brand analysis

extraction of implicit information will be done using various techniques (see Sect. 4). The result of the annotation phase is then transformed and loaded into a data warehouse (see Sect. 5). The ETL (extraction, transformation, and loading) process includes data transformation, cleaning, and aggregation, with the ultimate aim of computing various brand analysis metrics, which can be used for further analysis.

4 Data Annotation

During the annotation phase, we start from the collected data and we annotate them with brand, time, location, and sentiment information. Brand and time information can be obtained easily through standard keyword search. On the other hand, the analytical techniques employed to generate location and sentiment annotations are often specific to the type of data, and described in the following.

Sentiment Annotation. Sentiment annotation of Amazon reviews relies on readily available ratings (1,2 is negative; 3 is neutral; 4,5 is positive). For tweets and news, we rely on both natural language processing (NLP) and machine learning approaches.

For NLP, we considered SENTIWORDNET [1], which is a sentiment analysis extension of WORDNET, associating each WORDNET synset with three numerical scores $Obj(s)$, $Pos(s)$ and $Neg(s)$, describing how objective, positive, and negative the terms contained in the synset are. As an alternative approach, we also considered Google NLP,[1] a built-in service in Google cloud, which also includes sentiment analysis. Sentiment analysis attempts to determine the overall attitude (positive or negative) expressed within the text, representing it with a score ranging between –1.0 (negative) and 1.0 (positive). This score corresponds to the overall emotional leaning of the text. Further, a magnitude value between

[1] http://cloud.google.com/natural-language/docs.

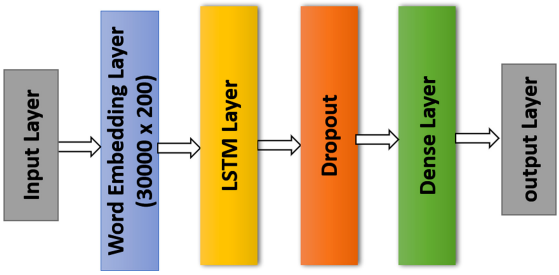

Fig. 2. Architecture of LSTM-NN

0.0 and $+\infty$ is provided, indicating the overall strength of emotion (both positive and negative) within the given text (longer texts may have greater magnitudes).

Regarding machine learning approaches, we first implemented basic classifiers, such as random forest, logistic regression, and KNN, by using SK-learn[2]. We trained the models on a corpus of tweets obtained by merging more than 5000 labeled tweets[3] together with about 1.6 million tweets[4]. To the best of our knowledge, no sentiment-based labeled news are available as ground truth. Hence, we considered only this tweet dataset as ground-truth for machine learning model training. Unfortunately, classical machine learning approaches failed to generate sufficiently accurate models in our experiments. Therefore, we decided to employ a deep learning (DL) approach for increasing the accuracy of the model. Additionally, deep learning approaches can take into account the word context, by building the meaning word after word, in an incremental way. To this aim, we rely on a Long Short Term Memory (LSTM) network, a special kind of Recurrent Neural Network, able to account for dependencies on different time scales, typical of natural language texts (see Fig. 2). We trained the model on the same dataset selected for the basic classifiers, using the Keras library[5]. The first layer (*input layer*) embeds each atomic word into a real-valued vector in a predefined vector space. Such embedding can be learned from large, unlabelled corpora using neural networks, and encode both syntactic and semantic properties of words [10]. Studies have found the learned word vectors to capture linguistic regularities and to collapse similar words into groups [17,18]. Their utility in tasks such as sentiment classification is well attested [12]. Based on these considerations, we trained Word2Vec [8] on about 30000 words from a tweets corpus (*word embedding layer*). As word embeddings alone have shown good performance in various classification tasks, we also use them in isolation, with varying dimensions, in our experiment. In the case of LSTM, a word embedding size of 300 resulted in high accuracy on the data set.

[2] http://scikit-learn.org/stable/.
[3] https://github.com/zfz/twitter_corpus/blob/master/full-corpus.csv.
[4] http://www.kaggle.com/kazanova/sentiment140.
[5] http://keras.io/.

To avoid the model to be over-fitted, we deactivate some nodes in LSTM, using a *dropout layer*. The fifth layer is a *dense layer* applying a linear operation, in which every input is connected to every output by a weight, followed by a non-linear activation function. In our case we use the sigmoid function to compute the probability (between 0 and 1) for a given text to be characterized by a positive sentiment. In the final *output layer*, the sentiment probability, computed by the dense layer, is transformed into a sentiment label as follows: the label is *negative* if the sentiment score s is $s \leq 0.4$, *neutral* if $0.4 < s < 0.7$, and *positive* if $s \geq 0.7$.

Location Annotation. For geolocating tweets, we rely on Geoloc [11], a recent state of the art data-driven geolocation algorithm, previously used by our group as a baseline technique for the definition of a knowledge-based geolocation approach [25][6]. Geoloc discretizes the Earth's surface into square cells of fixed size and models geolocation as a classification task, based on a kernel density estimation approach, with the aim of associating with each message a position corresponding to the most appropriate cell. For training this model, we considered a set of 4000 tweets, collected from and already annotated by Twitter, in addition to the training set of almost 5 million tweets provided with the model. Geoloc takes a tweet as input and provides coordinates as output. Coordinates are then enriched by considering knowledge-based information like city and country, using GeoSPARQL[7] queries on specific crowdsourced geographic ontologies, such as Geonames[8] and Open Street map (OSM)[9].

For news data, only few (and old) geolocated datasets are available (see, e.g., [13,15]). Hence, we developed our own approach for extracting geographic toponyms from the considered news dataset in two rounds. First, we extract entities cited in the text by relying on the Google entity analysis cloud service[10]. Then, based on DBPedia (an open linked dataset)[11], for each identified country, we add the capital as the city, and for each identified city, we add the corresponding country. If no location entity is found inside the news content, we follow the same process used for tweet annotation. Thus, we apply the Geoloc model to extract the coordinates of the text, then enrich those coordinates by applying geographic ontologies such as Geonames and OSM.

5 A Data Warehouse for Brand Loyalty Analysis

In order to analyze annotated data, we designed an ad hoc data warehouse. The data warehouse includes two data marts, one related to the *stock market*, providing stock market measures for brands at a given date, and one related to the *brand loyalty*. The resulting data warehousing schema, represented according to the Dimensional Fact Model [9], is shown in Fig. 3.

[6] http://code.google.com/archive/p/geoloc-kde/.
[7] http://www.geosparql.org/.
[8] http://www.geonames.org/ontology/.
[9] http://wiki.osmfoundation.org/.
[10] http://cloud.google.com/natural-language/docs/analyzing-entities.
[11] http://wiki.dbpedia.org/.

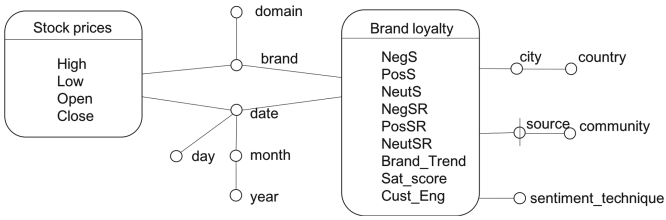

Fig. 3. Dimensional fact schema

Brand Loyalty Data Mart. Each fact in this data mart corresponds to customer opinions, computed by using a given *technique*, related to a given *brand* (generalized in the reference business *domain*), at a given *date*, in a given *city* (generalized in the reference *country*), and collected from a given *community*. The community represents the media from which texts have been collected (Amazon product reviews, Twitter, and news). Only for news, the community can be specialized into the specific data *source* (e.g., BBC, CNN news).

Each combination of the dimensional attributes leads to the identification of a set of texts, say A_i, annotated according to the considered technique. Customer opinions with respect to A_i are represented in terms of measures defined starting from metrics proposed for analyzing brand popularity and behavior in the market and, in particular, on key performance indicators (KPI) for brand analysis on social media [26], referring to the following groups: *Customer Experience (CE)*, measuring the loyalty on the sentiment basis [20,21,28]; *Customer Satisfaction (CS)*, focusing on how much the brand services satisfy the customers [3,26]; *Customer Interaction (CI)*, mainly dealing with the activity level of customers in taking on responsibilities for the company (e.g., elevating the company reputation by writing positive experience about one product) [4].

For what concerns the Customer Experience (CE), we consider three metrics: (i) the number of texts characterized by an either negative ($NegS$), positive ($PosS$), or neutral ($NeutS$) sentiment; (ii) the sentiment ratio for each sentiment type, $NegSR$, $PosSR$, and $NeutSR$, with respect to the total number of texts in A_i; (iii) the Brand Trend, defined as the ratio between A_i and the total number of annotated texts. For what concerns Customer Satisfaction (CS), when A_i corresponds to reviews, we compute the satisfaction score Sat_score as the average rating values assigned by the reviews in A_i.

Finally, when A_i corresponds to tweets, Customer Interaction is measured in terms of customer engagement $Cust\_Eng$, defined as the sum of the number of quoted status, retweets, and favourite selections for each tweet t in A_i.

Stock Prices Data Mart. This data mart provides stock market measures for brands at a given date. Measures correspond to the opening, the final, the highest, the lowest stock price for a brand in a given day. The two data marts together allow the analysis of the impact of customer opinions on stock changes.

Table 1. Sentiment approach comparison per brand

Brand	DL_WordEmb_LSTM	Google NLP	WordNetLexicon
Apple	0.52	0.59	0.53
Huawei	0.58	0.51	0.53
Samsung	0.56	0.59	0.54

6 Experimental Results

In order to discuss the flexibility of the proposed approach for brand loyalty analysis, in the following we first discuss how the pipeline has been customized for the analysis of data related to three important brands in the ICT domain, then we present some preliminary results we have obtained as examples of the analysis that can be performed.

6.1 Experimental Setup

In order to demonstrate the effectiveness of the proposed pipeline, we customized it for the analysis of the loyalty with respect to three ICT brands, namely Huawei, Samsung, and Apple. To this aim, we mined data from three communities, namely, Amazon product reviews, Twitter, and news articles, in the time window between February and May 2019, filtering textual clips with respect to the selected brand names.

For the news, we relied on the News API,[12] a simple REST API for retrieving articles from across the Web. We retrieve the "Top Headlines" news (breaking news) as well as "Everything", which is a firehose of news articles published by over 30000 news sources and blogs. To avoid a too high set of neutral classifications, often generated when the annotation is applied over long and generic texts like full news, we annotated only the summary of each news (up to 250 characters, also corresponding to the limit of the free access account) and we restricted the search to the 'Technology' topic (thus, avoiding, e.g., to collect news dealing with 'apple' as a fruit instead of 'apple' as a brand).

For Amazon product reviews, we developed a custom crawler that mines reviews related to products related to the considered brands. For Twitter, we collected the reference tweets using the Streaming API[13]. Finally, stock price data have been collected using the Yahoo Finance API[14]. Table 2 summarizes the volume of data collected for each community.

Our implementation follows a typical Extraction-Transformation-Loading (ETL) approach: we extract data from the sources as described above, then annotate and transform them using the algorithms and tools discussed in Sect. 4, then compute the measures described in Sect. 5, and finally store them into the

[12] https://newsapi.org/.
[13] https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data
[14] https://pypi.org/project/yahoo-finance/.

data warehouse implemented using HDFS, the Hadoop File System.[15] Brand loyalty analysis has then been performed by relying on Spark SQL[16] for analytical processing and Tableau[17] for data visualization.

Table 2. Brand data volume per community

Community	Apple	Huawei	Samsung
Twitter	2,371,344	1,143,697	1,247,109
Amazon	12,875	13,841	12,390
News	6,000	6,474	6,165

As a preliminary study, we compare the sentiment values generated by the three sentiment analysis approaches presented in Sect. 4 (Google's NLP service, WordNetLexicon, and the designed deep learning approach), after normalizing results in the range [0;1] (indeed, the deep learning approach delivers sentiment values in [0;1], with 0.5 referring to a neutral sentiment, the other two approaches deliver results in [–1;1], with 0 representing a neutral sentiment). The three algorithms have been applied to Twitter and news data. For Amazon reviews, as pointed out in Sect. 4, sentiment is extracted through explicit feedback given in the rating, so we excluded this community from this experiment. Average sentiment values computed by each technique for each brand are shown in Table 1. Although there are minor differences in the resulting sentiment value, we do not observe any significant difference nor does any approach consistently judge the sentiment of any brand too positive or too negative (values are around 0.5, thus they all refer to a neutral polarity). Hence, in the following, we rely on the designed deep learning approach for further analysis.

6.2 Results

Brand Sentiment Analysis. As a first experiment, we analysed *Customer Experience* by considering, for each brand, the average number of texts that, on a daily basis, have been classified with a negative, positive, or neutral sentiment (obtained through the aggregation of $NegS$, $PosS$, and $NeutS$ measures), to gain a perspective about the overall customer opinion for each brand through all communities. Table 3 shows that the sentiment value is mostly neutral for all the brands. Neutral sentiment statements indicate a quasi-objective mention of the brand, carrying neither overly positive nor negative emotion. Interestingly, there are substantially more positive expressions than negative ones for all brands. The most controversial brand in the study is Apple, for which we got an almost similar daily number of positive and negative clips. The number of positive opinions

[15] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
[16] https://spark.apache.org/sql/.
[17] http://www.tableau.com.

for Samsung is similar to that obtained from Apple, but, for the considered period, it attracts considerably less criticism overall. Table 4 refines the sentiment analysis taking into account communities. We observe that the highest sentiment values refer to Twitter for all brands. This is unsurprising, as this social media platform is often used to "vent". Another interesting observation is that there are considerably more negative sentiments about Apple in news articles than about the other two brands. This indicates that Apple might have an image problem in this specific community. Similarly, we observe that Apple has the least positive and the most negative sentiments also in Amazon reviews, indicating that some customers may be more dissatisfied with the related products than with respect to those of the competitors. These results demonstrate that it is crucial for SBI to evaluate brand opinion across multiple communities, as brand image varies, sometimes drastically, between them.

For what concerns *Customer Engagement*, Table 5 shows the average daily customer engagement on Twitter. Apple has by far the highest customer engagement across all the brands, followed by Samsung and Huawei, providing further evidence to the Apple brand being in general more polarising. However, note that Twitter is blocked in China (the homeland of Huawei). Based on the annual report of Huawei for 2018, more than 51% of their customers are from China. This might explain the relatively low Twitter engagement for this popular brand.

Finally, we compare the brands with respect to the average daily *Customer Satisfaction Score*, computed over Amazon reviews. Table 6 shows that Huawei has the highest satisfaction score, whereas customer satisfaction for Apple and Samsung is comparable.

Table 3. Customer Experience, by brand

Brand	Neutral	Avg sentiment negative	Positive
Huawei	10.36	4.12	8.31
Samsung	9.08	3.86	9.95
Apple	14.26	9.55	10.18

Geographic Analysis. Geographic information can help in further understanding *Customer Experience*. Table 7 points out the countries with the highest number of neutral, positive, and negative statements for each brand. As expected, the highest distribution of positive feedback for Apple is in the US, while the most negative sentiment is expressed in the Philippines. As for Samsung, both the most positive and neutral sentiments are expressed in Germany, while Benin (an African country) expresses the most negative sentiment about this brand.

Results about location-based *Customer Engagement* are shown in Table 8, reporting the countries with the highest Twitter engagement per brand. The results are in line with those obtained in Table 7 regarding Samsung and Apple. Thus, the highest customer engagement for Samsung is in Germany, while Apple

Table 4. Customer experience, by community and brand

Community	Brand	Neutral	Avg sentiment negative	Positive
News	Apple	1.45	2.93	2.72
News	Huawei	1.03	1.73	1.75
News	Samsung	1.03	1.83	3.37
Twitter	Apple	17.11	11.36	12.06
Twitter	Huawei	13.56	5.30	10.58
Twitter	Samsung	10.82	4.52	11.62
Amazon	Apple	0.07	0.39	0.71
Amazon	Huawei	0.08	0.20	1.00
Amazon	Samsung	0.09	0.35	1.10

Table 5. Customer engagement

Brand	Avg customer engagement
Apple	1,937,852.45
Huawei	156,962.065
Samsung	174,166.98

Table 6. Customer satisfaction

Brand	Avg customer satisfaction score
Apple	3.37
Huawei	4.049
Samsung	3.76

as the highest engagement in the US and UK. For Huawei, the highest engagement rate is in Switzerland.

Table 7. Location-based customer experience

Brand	Location-based sentiment		
	Positive	Negative	Neutral
Apple	United States	Philippines	United Kingdom
Huawei	United Kingdom	Poland	Spain
Samsung	Germany	Benin	Germany

Dashboarding. Using Tableau, we generated some SBI dashboards for brand analysis. As an example, Fig. 4 shows the most popular brands in EU countries while Fig. 5 shows brand sentiment values and changes in stock prices for all the three brands over the reference time period. We observe that for Huawei and Samsung, both metrics remained relatively stable. For Apple, the customer experience is stable over the time while the stock trend is up-wording till April then start to be stable.

Table 8. Location-based customer engagement

Brand	Countries with the highest Cust. Eng.
Apple	United States/United Kingdom
Huawei	Switzerland
Samsung	Germany

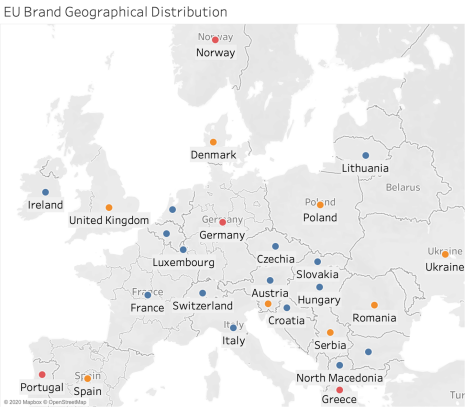

Fig. 4. Geographical brand strength

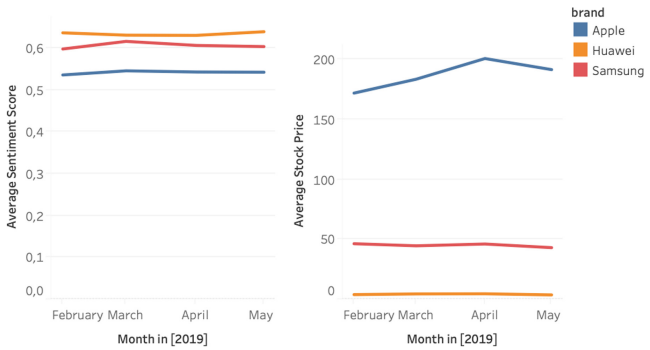

Fig. 5. Long-Term comparison of customer experience and stock prices

A second type of dashboard, comparing stock prices and customer engagement on Twitter, is shown in Fig. 6. The low direct correlation between them indicates that the effects of increasing customer engagement on stock prices may be indirect, delayed, or there may simply not be a direct impact.

Finally, Fig. 7 shows a dashboard for understanding the change of sentiment per community. The three line charts show that news sentiment is fluctuating the most (more evident for Apple and Samsung, somehow stable for Huawei). This is arguably because news coverage tends to be very sensitive to market

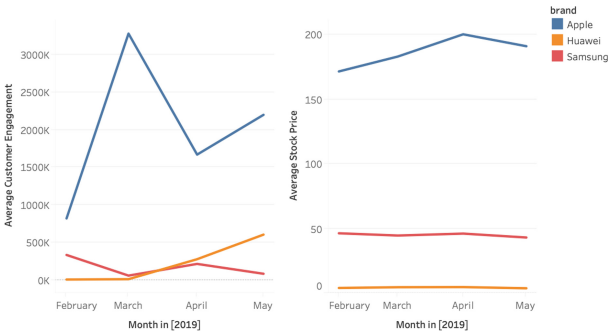

Fig. 6. Customer engagement and stock price trends

and policy events, as exemplified by the conflict between China and the USA. Twitter sentiment, on the other hand, is remarkably stable over time. Amazon reviews are rather stable for Huawei, but vary for the other brands. This indicates varying product quality for these brands.

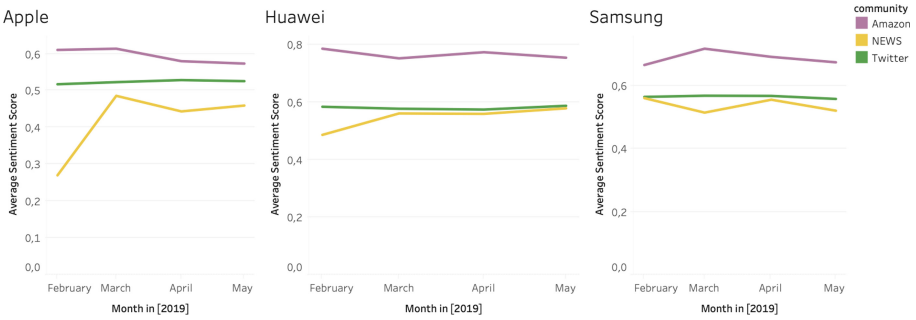

Fig. 7. Brands per community

7 Concluding Remarks

In this paper, we presented an SBI pipeline for brand analysis, relying on customer feedback gathered from social media. Data have first been collected from Twitter, news articles, and Amazon reviews, and then annotated with respect to sentiment and geolocation, relying on alternative approaches. Annotated data are then stored in a data warehouse for further analysis, through the usage of a specific set of measures tailored to brand loyalty analysis. We demonstrated the utility of the proposed pipeline through various experiments for three ICT brands (Huawei, Apple, and Samsung). The obtained (preliminary) results show the usefulness of considering a multitude of data sources and locations when analysing brand loyalty.

We remark that the presented results are just some examples of what can be obtained with the proposed pipeline and they have not to be taken as generally valid assessments about the three considered brands. For example, we investigated only three possible data source types (Twitter, Amazon reviews, and news articles). While we argue that the chosen sources are representative for common sources of brand information on the Web, this is evidently not a complete list of possible data sources. Our study makes no claims that our experiments will generalize to other sources of sentiment, such as software reviews posted on app stores or online forums such as Reddit. A similar limitation is related to the validity of the geographic analysis, due to the geographical availability of Twitter data. As already raised in Sect. 6, Twitter is currently banned in China, leading to the fact that the Huawei brand is not fairly represented in our Twitter data. Finally, we remark that different communities, locations, and policy events can impact all metrics and simple correlations are often not easy to find. Despite the issues discussed above, given that the main purpose of our experimental evaluation was to serve as an utility demonstration of our approach, we do not see these limitations as threatening the value of the conducted research as a whole: the proposed pipeline gives practitioners a framework to define their own metrics and analyse data for brands and communities of their interest.

In terms of future work, as a consequence of what stated above, additional experiments on more brands and more communities are needed to validate the proposed approach. Another important issue concerns the comparison of different KPIs with respect to their effectiveness in measuring brand loyalty, with the help of domain experts, also taking into account specific topic-related issues like, e.g., innovation and price policy.

References

1. Baccianella, S. et al.: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the International Conference on Language Resources and Evaluation, LREC (2010)
2. Castano, S., et al.: SABINE: a multi-purpose dataset of semantically-annotated social content. In: Vrandečić, D. (ed.) ISWC 2018. LNCS, vol. 11137, pp. 70–85. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_5
3. Cronin, J., et al.: Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. J. Retail. **76**(2), 193–218 (2000)
4. Kellogg, L.D. et al.: On the relationship between customer participation and satisfaction: Two frameworks. International Journal of Service Industry Management (1997)
5. Eisenstein, J. et al.: A latent variable model for geographic lexical variation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP (2010)
6. Gallinucci, E., et al.: Advanced topic modeling for social business intelligence. Inf. Syst. **53**, 87–106 (2015)
7. Gelernter, J. et al.: Automatic gazetteer enrichment with user-geocoded data. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, GEOCROWD (2013)

8. Goldberg, Y., Levy, O.: Word2vec explained: deriving Mikolov et al'.s negative-sampling word-embedding method (2014)
9. Golfarelli, M., Rizzi, S.: Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill, New York (2009)
10. Guggilla, C. et al.: CNN-and LSTM-based claim classification in online user comments. In: Proceedings of the 26th International Conference on Computational Linguistics, COLING (2016)
11. Hulden, M. et al.: Kernel density estimation for text-based geolocation. In: Proceedings of the 29th Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 145–150 (2015)
12. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the International Conference on Conference on Empirical Methods in Natural Language Processing, EMNLP, pp. 1746–1751 (2014)
13. Leidner, J.L.: An evaluation dataset for the toponym resolution task. Comput. Environ. Urban Syst. **30**(4), 400–417 (2006)
14. Lee, J., et al.: Integrating structured data and text: a multi-dimensional approach. In: Proceedings of the IEEE International Symposium on Information Technology, ITCC, pp. 264–271 (2000)
15. Lieberman, M.D., Samet, H., Sankaranarayanan, J.: Geotagging with local lexicons to build indexes for textually-specified spatial data. In: Proceedings of of the International Conference on Data Engineering, ICDE, pp. 201–212 (2010)
16. Liu, B. and Zhang, L.: A survey of opinion mining and sentiment analysis. In: Aggarwal, C.C., Zhai, C. (eds.) Mining Text Data, pp. 415–463. Springer, Boston (2012) https://doi.org/10.1007/978-1-4614-3223-4_13
17. Mikolov, T. et al.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 27th Annual Conference on Neural Information Processing Systems, pp. 3111–3119 (2013)
18. Mikolov, T. et al.: Linguistic regularities in continuous space word representations. In: Proceedings of Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, pp. 746–751 (2013)
19. Mudinas, A. et al.: Combining lexicon and learning based approaches for concept-level sentiment analysis. Association for Computing Machinery (2012)
20. Parasuraman, P.A., et al.: A multiple- item scale for measuring consumer perceptions of service quality. J. Retail. **64**(1), 12 (1988)
21. Parasuraman, A., et al.: A conceptual model of service quality and its implications for future research. J. Mark. **49**(4), 41–50 (1985)
22. Ravat, F. et al.: Top_keyword: an aggregation function for textual document OLAP. In: Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery, DaWaK, pp. 55–64 (2008)
23. Rahimi, A., Cohn, T., Baldwin, T.: Pigeo: a python geotagging tool. In: Proceedings of the ACL (System Demonstrations), pp. 127–132 (2016)
24. Rehman, N.U. et al.: Building a data warehouse for twitter stream exploration. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, ASONAM, pp. 1341–1348 (2012)
25. Di Rocco, L. et al.: The role of geographic knowledge in sub-city level geolocation. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC, pp. 687–689 (2019)
26. Stich, V. et Al: Social media analytics in customer service: a literature overview - an overview of literature and metrics regarding social media analysis in customer service. SciTePress (2015)

27. Suresh, S., S, G.R.T., Gopinath, V.: VoC-DL: revisiting voice of customer using deep learning. In: AAAI Press (2018)
28. Verhoef, P., et al.: Customer experience creation: determinants, dynamics and management strategies. J. Retail. **85**(1), 31–41 (2009)
29. Zhang, W., Gelernter, J.: Geocoding location expressions in twitter messages: a preference learning method. J. Spat. Inf. Sci. **9**(1), 37–70 (2014)