Operational Research (2022) 22:6007–6036
https://doi.org/10.1007/s12351-022-00714-0

ORIGINAL PAPER

Unsupervised consumer intention and sentiment mining from microblogging data as a business intelligence tool

Symeon Symeonidis[1] · **Georgios Peikos**[2] · **Avi Arampatzis**[1]

Received: 13 December 2021 / Accepted: 16 April 2022 / Published online: 12 May 2022

Abstract

The present study aims to create a framework that analyses user posts related to a product of interest on social networking platforms. More precisely, by applying information mining techniques, posts are categorised according to the intention they express, the sentiment polarisation, and the type of opinion. The model operates based on linguistic rules, machine learning, and combinations. Six different methodologies are implemented to extract intent, sentiment, and type of opinion from a tweet. The final model automatically detects intention to buy or not to buy the product, intention to compare the product with other competitors, and finally, intention to search for information about the product. It then categorises the text according to the sentiment and depending on their expressed opinion. The dataset comprises tweets for each day of the iPhone 5's life cycle, corresponding to 365 days. Additionally, it demonstrated that the business's external or internal decisions affect the public purchasing audience's opinions, sentiments, and intentions expressed on social media. Lastly, as a Business Intelligence tool, the framework recognises and analyses these points, which contribute substantially to the company's decision-making through the findings.

Keywords Intention mining · Sentiment analysis · Business intelligence · Microblogging

✉ Symeon Symeonidis
ssymeoni@ee.duth.gr

Georgios Peikos
g.peikos@campus.unimib.it

Avi Arampatzis
avi@ee.duth.gr

[1] Database and Information Retrieval research unit, Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi 67100, Greece

[2] Information and Knowledge Representation, Retrieval, and Reasoning (IKR3) Lab, Department of Informatics, Systems, and Communication (DISCo), University of Milano-Bicocca, Milan, Italy

1 Introduction

Nowadays, many companies, small, medium, or large, change their philosophy, bringing it closer to the consumer. This general shift leads to the development of new techniques that consider public opinion. Therefore, businesses need to identify a way that allows them to effectively consider customers' wishes and satisfy them somehow.

Moreover, in the last fifteen years, social media have been the most potent and effective channel for customers to voice their ideas as they are easy to use and flexible. Microblogging services have become a popular venue for individuals to express their intentions and wishes. Also, due to the extensive use of social media, such as Twitter, the public opinion expressed through such platforms contains helpful information regarding products and services.

In addition, technological advancements have led to a significant increase in online purchases and the ability to express reviews/evaluations via social media platforms. Specifically, using evaluations on social media due to the quick advancement of information technology. By using social media platforms, such as Twitter and Facebook, users can freely express their opinions and share their ideas, thoughts, experiences, and feelings (Shukri et al. 2015; Felt 2016; Wang et al. 2017; Poecze et al. 2018; Choi et al. 2020).

The management of large amounts of data coming from both internal and external business environments is beneficial for economic and financial analysis which are taken into account during a company's management decisions. In this unstructured information era, many valuable data related to the business activities are linked to their products and services. Concerning user-generated content (UGC) (Blythe and Cairns 2009; Smith et al. 2012; Wyrwoll 2014), businesses can utilize information related to their product or service by monitoring the related UGC during its life-cycle. As a result, in addition to their endogenous business data, a vast amount of free and easily accessible UGC can be employed in their decision-making processes (Zheng et al. 2012).

Currently, a number of studies (Hollerit et al. 2013; Purohit et al. 2015; Hamroun et al. 2016; Gao and Sebastiani 2016; Sun et al. 2018; Kurnia and Suharjito 2018; Kumar et al. 2019) are oriented towards the studying of texts related to services and products, also using data from social networks.

A motive for the proposed framework was to focus on utilized/collected UGC from social media related to a specific product or service and analyze it in terms of the expressed intentions, sentiments and opinions.

This study combines the recognition of users' sentiments, opinions, and intentions related to products and services through the extraction and analysis of data from social networks. Especially, we aim to analyse UGC from Twitter and identify four primary intentions:

(a) users' intention to purchase or not a product
(b) users' intention to compare a product with other competitors
(c) users' intention to seek information about a product.

Each of these intention-expressing tweets is further categorised simultaneously based on the user's opinion and sentiment. To conclude, the proposed novel framework, extract and analyse people's intentions expressed through social platforms during a product's life-cycle.

The framework's qualitative results are promising, demonstrating that the business's external or internal decisions affect public opinion, intentions, and sentiment expressed on social media. In other words, there is a feedback loop in place between businesses and users that allows for the extraction of critical information and use as a Business Intelligence tool.

The remainder of this paper is structured as follows. Section 3 provides an overview of the proposed framework, while its main components are further analyzed in Sect. 4. Section 5 describes the findings of our work. Finally, Sect. 6 concludes this research and outlines directions for future research work.

2 Related work

Several social media based techniques have offered to commercial companies or stakeholders numerous unique business prospects by monitoring client attitudes and determining their competitive position in the industry they operate. A consensus has been reached that social media data are appropriate for Business Intelligence research to understand consumers and markets better, while there is still debate about whether open data are preferred for BI research (Choi et al. 2020). Taking into consideration the following studies, we understand that users express in the same way their intention to buy related products. Businesses can build stronger relationships with their customers by listening and responding to specific customer comments. Also, UGC can help people think of online shopping as less risky before making a purchase decision Ladhari and Michaud (2015). Moreover unstructured textual big data is exceedingly difficult to manage for marketers, despite widespread agreement that UGC contains significant brand-related information Liu et al. (2017a).

2.1 Intention

In recent years, the extraction of user intent from the internet has become synonymous with user searches on various search engines. However, this tactic tends to be replaced by detecting user intentions through various forums and social networking platforms. Before analysing the theoretical framework in which each system's intention mining is implemented in this paper, it is considered appropriate to present a series of definitions of the concept of intention. Bagozzi (2010) presents a series of definitions related to the meaning of intention and the states that it is often found under the umbrella of the term will. A suitable definition of intent, adopted in our work, is the following: *Intention is a commitment, plan or decision of a person to carry out an action or to achieve a goal.*

An early research in the field of users' purchase intention mining is that of Ramanand et al. (2010), which placed purchase intention on the broader area of desires (wishes) while, at the same time, set a series of linguistic rules for finding it in texts that come from product forums. Later, Chen et al. (2013) reported that users express their intention to buy in a similar way for relevant products, such as mobile phones or televisions. Moreover, Gupta et al. (2014) examined the purchase intention in combination with the polarization of emotion for posts on social networks. To improve their model, they used a list of specific verbs that indicated purchase intention. Finally, due to the heterogeneity of the intention classes, they employed the ROC curve to evaluate the model, where AUC (Area Under the ROC Curve) reached 89%.

In mining intent to search for information about products and services on social media, Morris et al. (2010) conducted a study of how users seek information via search engines and social networks. Their research pointed out that most users search in search engines, but when they want to consult for specific products, they decide to post on a social network and receive personal answers they trust. In a related study, Paul et al. (2011) developed a set of 5000 data points after preprocessing a random sample of 1.2 million tweets to categorize the data points according to their topic matter. They conclude that 16% of the tweets are questions in the form of polls related to high-tech products. Finally, Zhao and Mei (2013) suggested that by analyzing the information needs of users on social media and specifically on Twitter, Google search trends can be conjectured. This fact proves the gradual passage from the established search form and the prevalence of a new direction. Jindal and Liu (2006) studied the detection of comparative proposals for products in texts using linguistic models, while the concept of product comparison intent is often defined as competitive ntelligence. In addition, Xu et al. (2011) focused on comparative mining suggestions from consumer reviews on forums and achieved significant results in categorizing proposals containing product comparisons. At the same time, they identified propositions, analyzed and deconstructed them into constituent elements.

2.2 Sentiment analysis

Sentiment analysis concerns the automated extraction of positive or negative opinions from texts and involves the computational detection and investigation of opinions, feelings, emotions, and subjectivities in texts (Li and Wu 2010; Liu 2010). Sentiment analysis is frequently used to examine large amounts of user-generated content (UGC) posted on social media and online shopping sites. Given the importance of online product reviews in influencing customers' purchasing decisions, many studies use sentiment analysis to investigate the advantages of online reviews or to rate competing items (Salehan and Kim 2016; Liu et al. 2017b). Furthermore, sentiment analysis can detect the polarity of sentiment from short and low-quality text, allowing social media data, which consist of short phrases with few keywords or mostly stopwords such as abbreviations and emoticons, to be analyzed almost accurately (Guzman and Maalej 2014). Lexicon-based approaches use lists of words annotated by polarity or polarity score to obtain the overall sentiment score of a given text. The key benefit of these strategies is that they do not need any training data (Giachanou and Crestani 2016; Kalamatianos et al. 2018). According to Hung and Lin (2013), existing lexicons are widely employed in lexicon-based studies because of their broad coverage of terms and avoidance of the dependability issues that plague the manual generation of sentiment dictionaries.

2.3 Business intelligence

Business Intelligence (BI) refers to decision support systems based on integrating and analysing organizational data resources to improve corporate decision-making. In addition, the term is often used to denote a range of various uses of data analysis that allow for more informed decision-making based on a larger body of knowledge (Watson and Wixom 2007). According to Negash and Gray (2003), BI describes the broad field which, by fusing structured and semi-structured data with warehouse management and data analysis techniques, can contribute to decision-making in order to reduce the time required, while at the same time help to improve quality of decisions.

In the last years, some research focused on the organizational effects of BI, claiming that implementing BI systems in an organization entails not just technology advancement but also a new manner of executing and controlling business operations and decision-making processes. Companies strive to earn and maximize profit via the selling of goods and services on a global scale. Therefore, organizations may use BI to systematically comprehend sales data and tailor trade strategies to clients' demands while attracting new customers and keeping old ones with value-added items.

Businesses can use BI to integrate powerful tools, analysis, standardized reporting, a monitoring system with various metrics, data integration, and other features into a service-oriented architecture, which is necessary for good business management to guide managers in strategic directions for quality information, as well as the establishment of standards and procedures, to ensure compliance with the objectives (Eckerson 2010). Businesses should use a robust and proactive approach, changing supply to consumers, investigating a more competitive pricing model, eventually replacing current markets with new markets, and becoming more competitive than competitors (Reeves and Deimler 2009).

Some studies have examined the relationship between social media analytics and business intelligence. For example, Rui et al. (2013) investigated the impact of tweets on movie sales. They discovered that the valence of social media data has a considerable effect on a consumer's propensity to view specific movies, leading to the conclusion that social media analysis should be utilized to estimate movie sales in the future. Furthermore, Sharma (2013) studied the life-cycle of rapidly evolving consumer products in the four stages that a product passes from its entry into the market until its withdrawal, while extensive reference was made to the strategies that the company must follow in each stage of a products' life-cycle. Finally, Wiecek-Janka et al. (2017) used a series of data from the official website of Apple and other sources, creating a graphical representation of the life-cycle of products of interest to present their life span and how they match with Apple's life-cycle.

Nevertheless, Tutunea and Rus (2012) pointed out that cost-effective cloud-based BI solutions exist and that an effective BI solution may provide Small and Medium-sized Enterprises (SMEs) a competitive advantage. However, most SMEs still have difficulties selecting and implementing an appropriate BI solution due to a lack of knowledge and technical know-how.

2.4 Weak supervision

In the last years, the limitation on domain-specific labelled datasets and the high cost of creating and labelling a dataset turned most professionals into weak supervision. Approaches such as distant supervision, where records from an external knowledge base are heuristically linked with data points to provide noisy labels (Alfonseca et al. 2012), and iterative rules and heuristics for labelling small subsets of data (Varma and Ré 2018), have tried to improve weak supervision methods. However, as a result, weak supervision has two main problems: poor performance due to noise in the label creation process and trouble construing the generated labels (Nashaat et al. 2018). The main differences between weak supervision and the proposed framework are mainly focused on the time it takes to classify a sufficient number of labels and the time spent on functions creation.

3 Proposed model

Social media analytics usually employ data mining, text and sentiment analysis methods to create new content, assist in decision-making, and improve competitive advantage (He et al. 2015). The purpose of this study is to determine whether social media platforms, notably Twitter, are suitable venues for businesses to extract user intentions. The target is to create a framework for mining user intentions related to a product during its life-cycle and using it as a BI tool for decision support.

In this respect, the first requirement was finding a specific Product of Interest (PoI). This product should have a clear, predefined life-cycle, determining the sampling period for crawling the data. Apple's iPhone 5 undeniably had a significant impact on the market; despite being the best-selling iPhone, it was withdrawn from the market just after a year. For this reason, in our study this smartphone model has been chosen as the PoI (Wiecek-Janka et al. 2017). Moreover, this PoI was selected to ensure that it is directly traceable in 1% of the Twitter stream.

The second requirement is related to the source of the data. Collecting a sufficient quantity of data containing public opinion about the (PoI) is crucial for this framework. In a related study, (Chen et al. 2013) focused on specific technological fora. Nevertheless, those sources of fora cannot provide sufficient day-to-day data about our research selected PoI and fail to offer domain relatedness since numerous data sources are merged as a single source domain. For this reason, Twitter was selected, to crawl its data and export insights containing UGC related to technology (Xu et al. 2017).

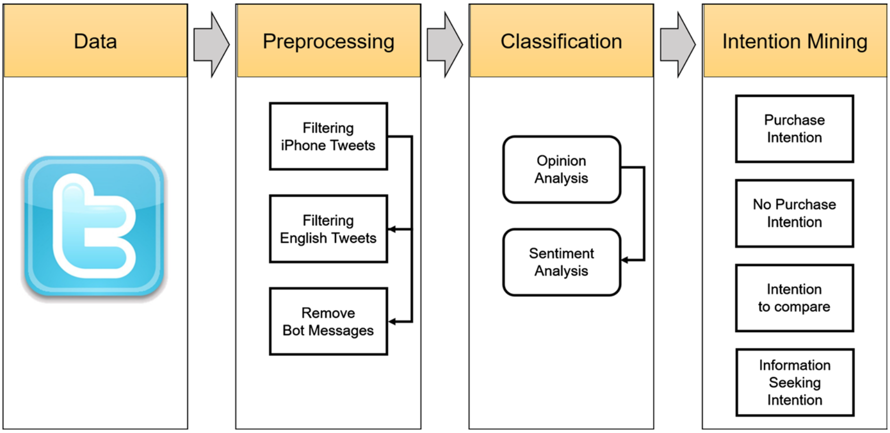

Fig. 1 Framework for mining intention

The proposed framework can search for specific user intentions related to the PoI in a large amount of data, fast and accurately. Specifically, theoretically argued and quantitatively proved that:

- Social networks, especially Twitter, are places where users exchange opinions about products, and therefore there can be helpful insights for a business hidden in the data (Vidya et al. 2015; Kim et al. 2016; Sun et al. 2018).
- Different filtering models can be used to extract user intentions related to a specific product in real-time and create a solid BI framework to help decision-making for a business.
- Internal and external market events (e.g. release of a competitive product in the market) related to the PoI and the business can be identified using Twitter. Tweets corresponding to these events are related to the specific intentions of a user and are expressed through the content of these tweets.
- Each intention studied is expected to align with other specific characteristics such as sentiment and opinion polarity. For instance, it is anticipated that negative sentiment outbalances the intention not to buy the product or objective opinion transcends the intention to compare.

Given the name of the PoI and tweets crawled during its life-cycle, in the end, this framework categorizes tweets into four categories according to user intentions. Also, the framework recognizes the sentiment polarization and the objectivity or subjectivity of each tweet.

Figure 1 presents the framework outlining our research methodology, from receiving the data to detecting the intention categories. Firstly is the collection of 1% of the Twitter flow for the complete life-cycle of the PoI. The next step is to export the tweets that contain the PoI in their text, preserve those classified as English based on the metadata field, and remove the bot-tweets ending up in the final data set. In the next step, the procedure for detecting opinion and sentiment is presented, which is applied to all English noise-free tweets. Finally, the intentions of the users are assigned to the specific categories. The following sections will present and explain in detail all the steps shown in the above figure.

Consumers actively express their thoughts online, thanks to the emergence of weblogs, online consumer fora, and product comparison sites. As a result, the majority of these consumer thoughts are now freely accessible on the internet. Furthermore, these are accessible to various industries, including financial services, telecommunications, and consumer products. Automated intention analysis based on such evaluations might be a cheaper and faster way to sense and uncover insights about a customer's thoughts and activities, complementing the more traditional survey approaches (Ramanand et al. 2010).

The following main concerns describe in what way the proposed framework relied on studying user intention on social networking platforms and extracting valuable information concerning a product.

(i) How will the data be collected so that they represent the public opinion and lead to safe conclusions?
(ii) How will the processing and final filtering be done, in order to reduce the amount of data?
(iii) How can a business be sure that the intentions' estimation and related characteristics are trustworthy?
(iv) What is a proper evaluation approach in this kind of problem?

4 Methodology

4.1 Data collection and pre-processing

In a business, products that comply with the rule of fast-changing consumer goods, such as mobiles, are the most difficult to manage in today's market (Faulds et al. 2018). When a product enters the market, it goes through predetermined stages, from its introduction to its withdrawal. For example, Apple's smartphones have a two-year average lifetime based on a model's release date, and new models are issued once a year. Therefore, within these twelve months, decisions must be made as accurately as possible to deal with other competing products and maximize profit (Wiecek-Janka et al. 2017).

The shorter the life-cycle of a product and the faster the transition from one phase to the next, the more important the need to optimize the decision-making within the business. Therefore, the model implemented in the present study may provide a decisive advantage for a business, specifically SMEs.

At this point, there are two possible ways of downloading the data. Initially, the data may introduce the PoI as a keyword in the Twitter API. This method will produce more product-related data but fails to obtain a complete picture of the appearance of the PoI in ratio to the total number of tweets in a given timespan.

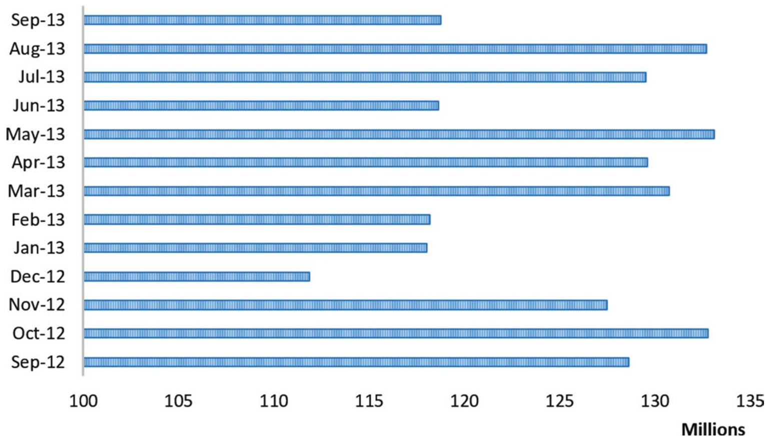

Fig. 2 Downloaded tweets per month

The second method is to get at least 1% of the Twitter Feed without the PoI as a keyword. In this case, the dataset is imprecise, but it gives a satisfactory statistical notion of the frequency of product-related tweet occurrences.

It is an argument that if the studied PoI was produced by a well-known and dominant product, a business could exploit both aforementioned approaches. In contrast, for SMEs, as their product is unlikely to be found in the 1% of the Twitter feed, the first approach is more reasonable. In any case, it is possible and reasonable to combine both of the approaches as mentioned above.

Wiecek-Janka et al. (2017) used data from the four-month sales of the Apple mobile products to present the life-cycle of this research proposed PoI and all models of the company. Therefore, the proposed framework relying on this state used the benchmark of four-month sales to determine the sampling interval of data from Twitter is fulfilled. It starts from September 21, 2012, to September 10, 2013.

The Internet archive[1] was used to access the data. An initial dataset was created by retrieving an average of 3,000 tweets per minute for every day of the product life-cycle, ending with about 1.629.981.000 tweets. As can be seen in Fig. 2, the sample is not consistent across months, and we have also collected all of the accessible metadata for each tweet.

After filtering only the tweets related to and mentioned to the PoI, in this experiment setup on iPhone 5, the set ended up with about 3.000.000 tweets. The final collection of data used to extract users' intentions consisted of 1.243.249 English tweets which referred to the PoI in their text, as shown in Fig. 3.

The collected Tweets, as mentioned above, were from the 1% of the daily Twitter stream and originated from a random sampling. Moreover, the data was distributed over the life-cycle of the PoI and was independent of the geographical area. It must

[1] https://archive.org/.

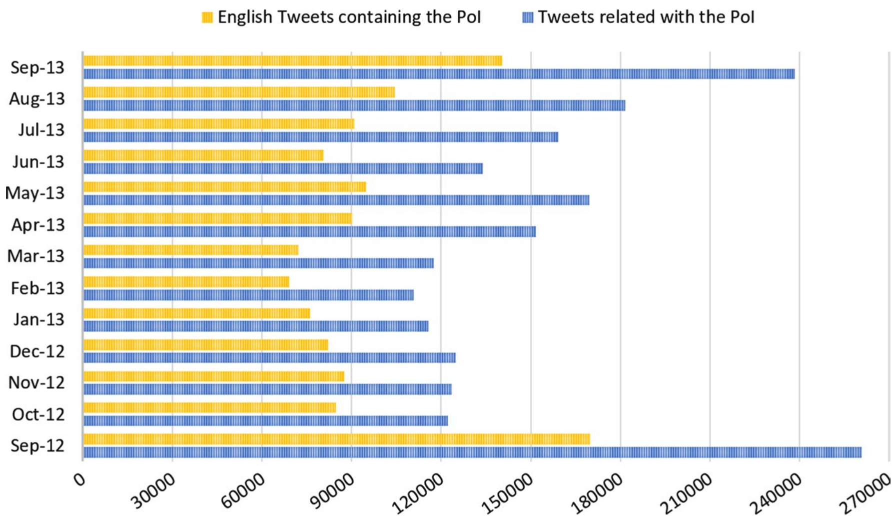

Fig. 3 Tweets related with the PoI and English Tweets containing the PoI

Table 1 Examples of malicious Tweets in the collection

Source	Tweet
The Tribez on iOS	I've collected 24,773 gold coins! http://t.co/Er3Of T6bbB #iphone, #iphonegames, #gameinsight
Runtastic	Just finished a runastic bike trip!!! I run 4.63 miles
Paradise Island: Exotic	Just Reached Level 20 on Paradise Island Fights!!

be stressed that the PoI is a topic that has a considerable influence on public opinion, in absolute frequency, although the percentage seems small.

4.2 Noise reduction and removal

Twitter data are widely used both by businesses and academia to retrieve valuable information and solve problems. When working with this kind of data, it is crucial to filter out the noise they contain (Symeonidis et al. 2018). In this current task, removing malicious tweets is vital as the PoI was given as a gift to many competitions or used as click bait by malicious websites. Therefore, those tweets were not a real representation of a user's interest for the PoI. Examples of malicious tweets in the collection are presented in Table 1.

The way malicious tweets were isolated came from a combination of two significant research works. The first is that of Lumezanu and Feamster (2012), which presents a comparative study of spam via email and social media posts. They stated that the existence of a URL in a tweet text was enough to categorize it as malicious. The study of Tsou et al. (2017) used Tweet metadata to classify them as malicious.

Table 2 List of Tweet sources; categorised as acceptable and improper

Good sources	Web, Twitter for iPhone, Twitter for Android, Twitter for BlackBerry, Twitter for Mac, Twitter for iPad
Bad sources	The Tribez for iPhone, The Tribez on iOS, My365, Tweetbot for Mac, novisa apps,Tweetbot for iOS

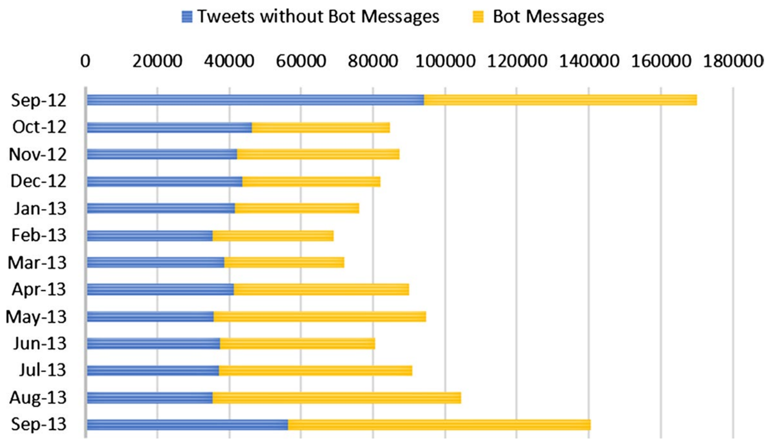

Fig. 4 Bot messages and final set of Tweets

This study's revealed source of spam or bots differed from prior published sources and were updated with these of Chatterjee and Krystyanczuk (2017).

Table 2 presents the lists of some of the good and malicious sources from Twitter posts, as our method automatically identified by our method.

In the process of filtering using a series of tweet metadata and the text of the tweets themselves, the majority of the noise (Bots, Spam, or Cyborgs) was successfully isolated. Finally, tweets that came from a good source and did not contain a URL were kept in. As it can be seen in Fig. 4, the majority of the obtained tweets were identified as noise (bots), proving the necessity of their removal.

Another way to deal with noise, which has been tried but rejected as the former was faster and better, was calculating the similarity of tweets with a set of Bots, Spam, or Cyborgs tweets used as root. However, the PoI chosen for the research dates back to 2013, when removing malicious tweets was relatively easy. Nowadays, more sophisticated approaches may be needed (Sahoo and Gupta 2019).

4.3 Purchase intention mining

Extracting users' intent is a challenging linguistic problem. Therefore, to derive users' purchase intention, we relied on the existing literature and created a set of linguistic rules. The process which extracts users' purchase and users' no purchase intentions can be seen in Fig. 5.

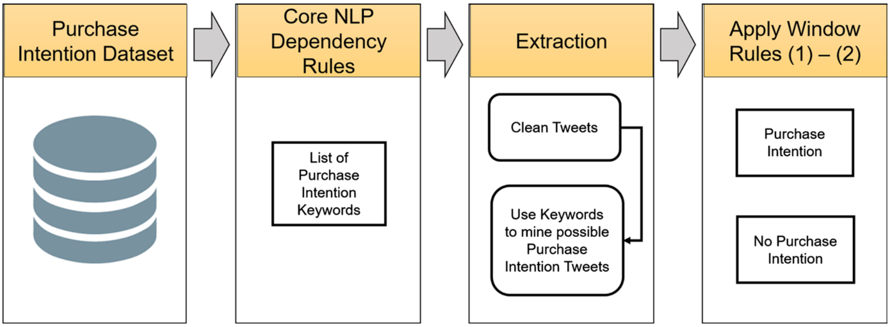

Fig. 5 Extraction methodology for purchase intention

Table 3 Purchase intention verbs and phrases

Verbs	Looking, get, buy, want, need, purchase, use, replace, choose, take, find, like, thinking
Phrases	Would like, should buy, would choose, could get, could find, would use, would have

Table 4 Examples of tweets with purchase intention

Tweet
I swear i want that new iphone
RT @USER: I need the iphone 5 in my life.
Finna get an iPhone this week.. yuppp
I m be getting da iPhone 5s beofe ya'lall

In the first stage, the Core NLP[2] library was used, to extract terms and phrases from an annotated dataset with forum data on product purchase intent (Chen et al. 2013). The first CoreNLP dependency used was the 'modal auxiliary', having as filter a series of modal verbs directly related to the expression of wishes, desires, and market intention (Ramanand et al. 2010). The second linguistic rule was to extract all the verbs in the class labeled as a purchase intention class. Table 3 shows the six verbs, and expressions with the highest frequency on the previous work of Chen et al. (2013). The final list of purchase intention keywords contained 20 terms with the highest frequency.

In the second stage, these terms or phrases were used within a three-word window along with the PoI to identify if the user expressed a purchase intention in a tweet. As we have seen, there is an overlap between different iPhone models at some periods. Therefore, the system rejects tweets that mention other iPhone models to avoid wrong conclusions, and we applied a rule to solve this problem by specifying

2 https://stanfordnlp.github.io/CoreNLP/.

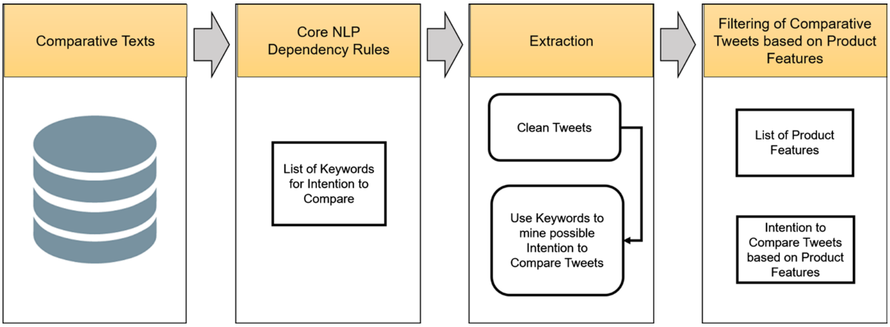

Fig. 6 Extraction methodology for compare intention

Table 5 Keywords to extract comparative user intentions

Keywords	Prefer, fewer, as easy as, comparable, as good as, one of the, higher, better, exceeds, compared to

that the PoI must be within a three-word window. Moreover, an additional linguistic rule was applied to detect and extract non-purchase intentions for the PoI. To accomplish this, the existence of any form of denial within the window containing the PoI and the keyword was considered enough. Examples of Tweets with Purchase Intention are presented in Table 4.

4.4 Comparative intention mining

Mining users' comparison intentions are one of the most critical features of our framework. Market conditions of products that are characterised by a short life-cycle, such as smartphones, require constant monitoring. Therefore, these data are essential for the fulfillment of this purpose (Jhamtani et al. 2015).

Motivated by previous studies (Carlos and Yalamanchi 2012; Chen et al. 2013; Ding et al. 2015), we created a new methodology for mining users' comparison intentions from social media platforms. The structure of our model can be seen in the Fig. 6.

Comparison of products can be analysed by identifying the following entities, according to Montoyo et al. (2012):

- Main Entity
- Secondary Entity
- Comparison Words / Phrases
- Common Characteristics of Entities

The most valuable part of this methodology is extracting appropriate words and phrases from a labeled dataset using syntactic dependencies. To this aim, Python's

Table 6 Examples of tweets with comparative intention

Tweet
I know I've tweeted this before, but I really need the iPhone 5 to come out.
My current phone is sad.
RT @USER: Need an iphone asap, cannot deal with this blackberry any longer.
@USER your commercials are so bad and your attempts to slander the iPhone make you look pathetic. Just accept 2nd place... At best.
I dont know why twitter on iPhone is much cooler than on android :p

Table 7 List of iPhones 5 comparative products

Comparative products	Blackberry, motorola, sony experia s, htc, nokia lumia, lg nexus, samsung

programming language library NLTK[3] along with a labeled dataset that contains product comparisons (Jindal and Liu 2006) was used.

The PoI is the critical element of this framework, and the sampling method ensures its presence in tweets. Using Stanford University's Core NLP library (Manning et al. 2014), and a series of complex linguistic rules, thirty-five (35) keywords and phrases came up to extract product comparison intention. The following Table 5 shows the ten most common keywords in the dataset.

After filtering the dataset with these keywords, tweets containing users' intentions to compare the PoI were extracted. Then, in the second step, the returned tweets were further processed to reduce their volume, so that the system could manage them in real-time. A list of features related to our PoI was employed to achieve that. Those features were used as an additional filtering level to allow the system to deliver better results to its end-user. For this reason, the procedure only returns a tweet if it contains any term comparison keyword and, at the same time, if there is a PoI feature mentioned in it. As the studied PoI was a mobile phone, words like camera, SIM card, GPS were related as keywords referring to a PoI feature. Examples of Tweets with comparative intention are presented in Table 6.

The secondary entity can be inserted into the framework as an extra input, but we chose not to use the competitors' list as an input for our system. By doing that, the framework can identify and analyze the possible interaction between the PoI and all the competitor products. In contrast, a specific PoI-competitor conclusion can be drawn when the end-user specifies the secondary entity. Table 7 provides a list of products, present during the PoI's life cycle, that are its competitors, identified in Yao and Sun (2016).

[3] https://www.nltk.org/.

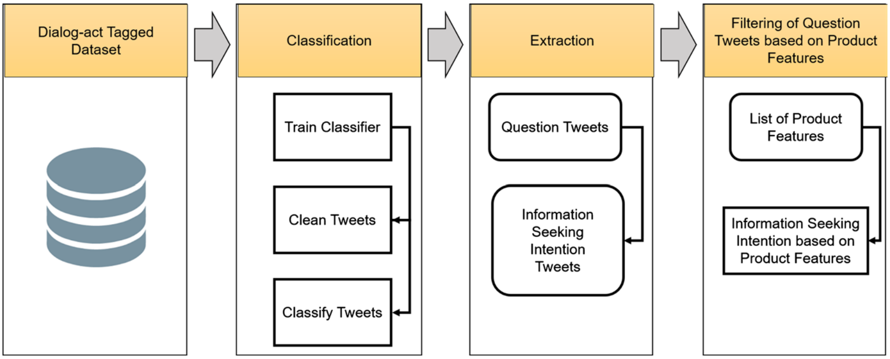

Fig. 7 Extraction methodology for seeking information intention

Table 8 List of iPhone 5 features

iPhone's 5 features	Price, camera, applications, iOS battery, screen, case, warranty

4.5 Information seeking intention mining

Social media are undoubtedly a modern way of communicating and exchanging views between people. For example, users often turn to Twitter to search for opinions and information about their buying plans. Therefore, in the data that we had at our disposal, there were questions from users about the PoI from which the framework can derive business insights.

According to Sharma (2013), market research is critical during the product's development phase to inform the company about which product features are appealing and should be avoided. Then, the company provides the forthcoming buyer with a product with convenient and valuable features, expedite the introduction phase and increase sales. Thus, our proposed framework contributes to that.

Extraction of users' intent to seek information on the PoI can be broken down into four stages, producing two sets of tweets. The first set looks at the content of tweets interspersed with many generic questions. Then, the second set eliminates the questions reached from the first set and focuses on particular product features. The used methods are shown in Fig. 7.

In the first stage, the Naive Bayes classifier was trained with the NPS Corpus (Jhamtani et al. 2015), which consists of 11,000 texts rated according to the dialogue-act tag (Stolcke et al. 2000; Wu et al. 2005). The algorithm accepts a text as input and produces the type of speech it represents. The outputs can be acceptance, clarification, emphasis, rhetoric, rejection, statement, question, Yes / No Question. From these outputs, tweets labelled as a question or Yes / No Question were kept, and the classifier had a 75% classification accuracy.

At this point, all extracted question tweets have been retrieved, and the framework can analyze them. In the next step, a list with specific PoI features was created

Table 9 Examples of Tweets with seeking information intention

Tweet
Looking for new iPhone cases.. Any suggestions?
Is iPhone 5 battery "long lasting" enough?
Why does the front camera of an iPhone have to be low quality?
So what's new with the iPhone 5?

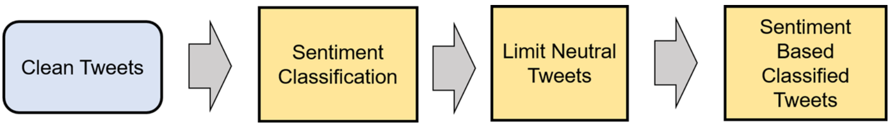

Fig. 8 Sentiment polarization detection

based on the literature review. We decided that this list will be an extra entry into the framework, depending on the PoI.

A simple linguistic rule was applied, demanding the presence of one of the PoI features listed below in Table 8.

Using the feature list part of which can be seen above, the volume of the data decreased significantly. Examples of Tweets with Seeking Information Intention are presented in Table 9.

4.6 Opinion and sentiment analysis

When users express their opinions, they frequently insert emotions into the message. Consequently, continuing with the previous assertion, extra analysis is required for a text where the user's intent is uncertain and where opinion is frequently detected.

Unlike the extraction of intentions that have recently received the scientific community's attention, the extraction of emotions and opinions is one of the first areas studied. There are plenty of Python libraries that can perform opinion and sentiment analysis.

In Fig. 8, the steps to detect the tweets' sentiment polarization using two different Python libraries are demonstrated, TextBlob[4] and VaderSentiment (Hutto and Gilbert 2014).

While working with these libraries, it is essential to tune the thresholds that categorize the text to neutral, positive, or negative. For this framework, the tuning of the parameters was done using a Twitter sentiment dataset obtained in 2013 (Saif et al. 2013). Therefore, using each library, an algorithm tested all the possible threshold combinations. Then, the best threshold values were chosen based on the accuracy of classifying tweets.

[4] https://textblob.readthedocs.io/en/dev/.

Table 10 Examples of Tweets with sentiment and opinion

Sentiment	Opinion	Tweet
Negative	Subjective	The only thing I don't like about my iPhone is how It always deletes my contacts.
Positive	Subjective	@USER I can't take this phone any more . And I like all the cases and the messaging app. Hopefully ill get the iPhone 5
Neutral	Objective	does anyone know if there's an untethered jailbreak for iOS 6.0.1 for the iphone 4???
Negative	Objective	it annoys me how i can't see iphone emoticons on google chrome

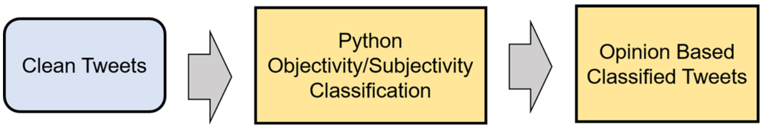

Fig. 9 Opinion classification

The two different libraries mentioned above were used to overcome the problem of neutral sentiment categorization of the tweets. Finally, the two algorithms were combined only in those tweets that the TextBlob algorithm classified as neutral. In this case, the final classification was done by the VaderSentiment algorithm. Examples of Tweets with Sentiment and Opinion are presented in Table 10.

Due to a lack of annotated datasets for the study period, the opinion threshold for classifying tweets (objective or subjective) could not be determined. The type of user opinion was classified based on TextBlob's library threshold as presented in Fig. 9.

5 Results

The implemented framework goes beyond sentiment and opinion analysis. It relies on many different methodologies of relevant research fields to study more complex problems such as the mining of consumers' intentions. This section aims to present the model's quantitative results when applied on real Twitter data.

Initially, in Fig. 10, the number of tweets was presented, classified in the studied intention categories. After preprocessing the data, judging from the results presented in the chart below, it is evident that consumer intentions studied in this work make up about 50% of the dataset.

One can observe that 7% of all tweets are classified as information-seeking intentions, from which only 1% relates to information-seeking based on specific PoI features. A similar observation is drawn from the category of the intention to compare the PoI with its competitors. Separation based on the characteristics of the PoI was made as it can help in the decision-making during different phases of the product's life-cycle.

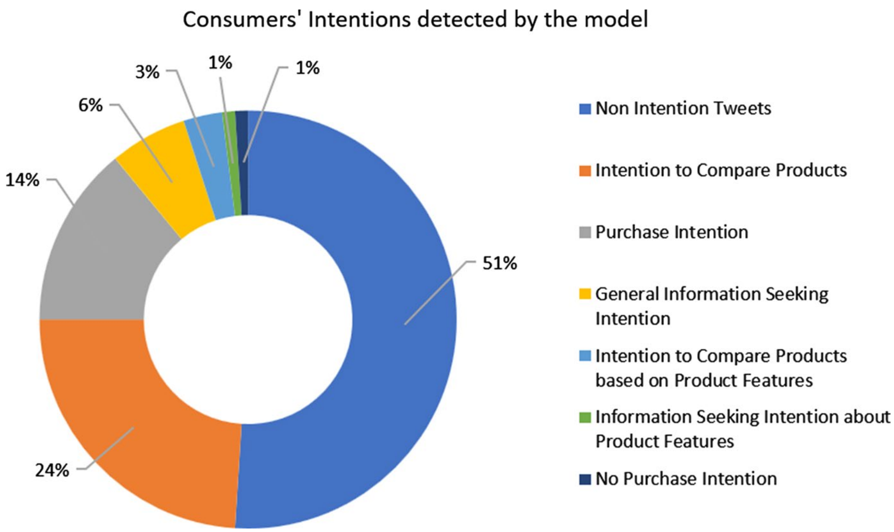

Fig. 10 Percentages of intentions of users in the data set

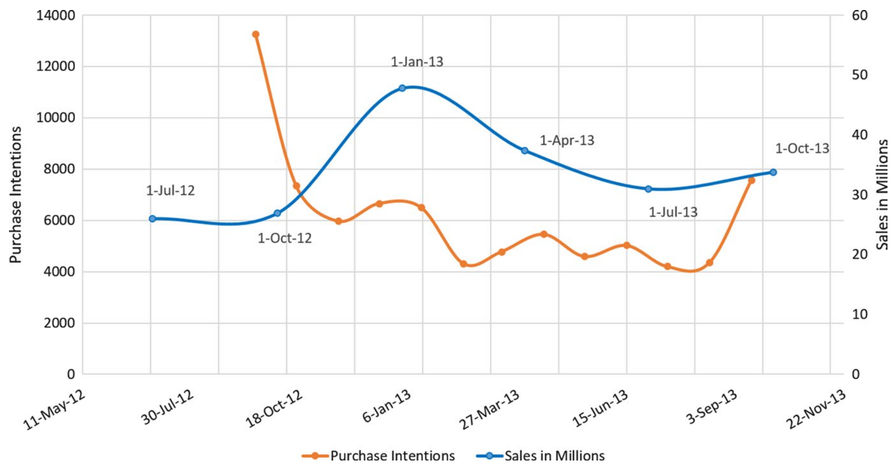

Fig. 11 Relationship between intentions of purchase and number of sales

Following, the results of each type of intention studied in this paper in more detail will be presented. Then, finally, every intention with the sentiment polarization and the kind of opinion will be combined.

5.1 Purchase intention results

Purchasing intent expressed by potential customers is the most valuable insight for the business as it is related to potential sales. An important strategy to ensure

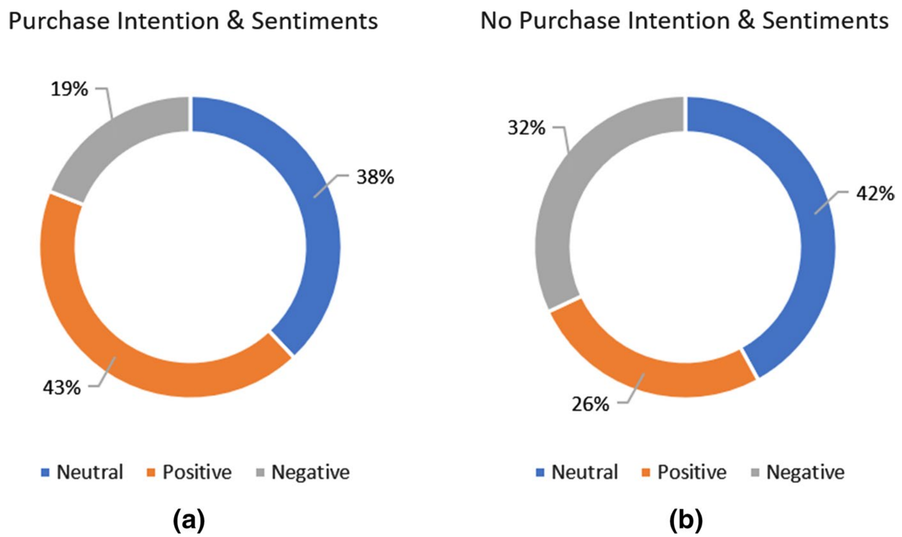

Fig. 12 Purchase intention and sentiment (**a**), No Purchase Intention and Sentiment (**b**)

higher profits is the proper distribution of the product upon its entry and withdrawal from the market. At this stage, the business has to select the stores to which the product will be shipped. Significant aid in this could be to determine the geographical distribution of users with purchase intentions.

Figure 11 presents the life-cycle of the iPhone 5, created from the number of sales obtained by Apple data. Due to the inability to separate sales of the iPhone 4s (which is not a product of interest) and the iPhone 5, purchase intent for both devices is presented in the same chart.

The model predicts users' purchase intentions before the actual increase in the sales for the company, and that proves that consumers often express their intent to buy a product via microblogging before buying it. At the maturity stage, from December 2012 to July 2013, of the product and afterwards, we notice that users' purchasing intentions decrease, but still remain at high levels. Therefore, one can conclude that the purchasing intent of the public has gradually declined, so the company has had a smooth phase of maturity. If the downturn had been more significant and steeper, perhaps the company should have considered applying a cannibalistic strategy to pull the model off the market and introduce a new one. Based on the curve of the PoI, which appears to have a steep knee of a curve reduction in the maturity phase, one can hypothesize that this is why the iPhone 5c and 5s were released earlier in the market. Noteworthy is a presentation of the relationship between users' purchase intentions and sentiment polarization. For example, in Fig. 12, and especially in Fig. 12a, most of the tweets belonging to the Purchase Intent class were observed are characterized by positive sentiment.

A conclusion is that users express themselves positively or neutrally when they indicate their intention to buy. However, that is not the case if we study the

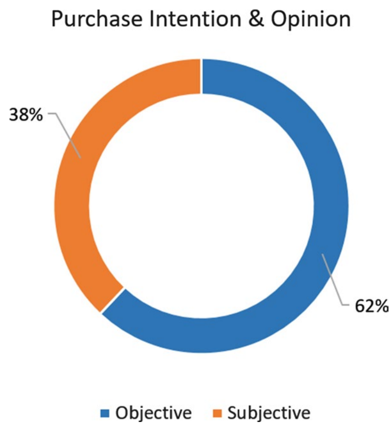

Fig. 13 Purchase intention and opinion

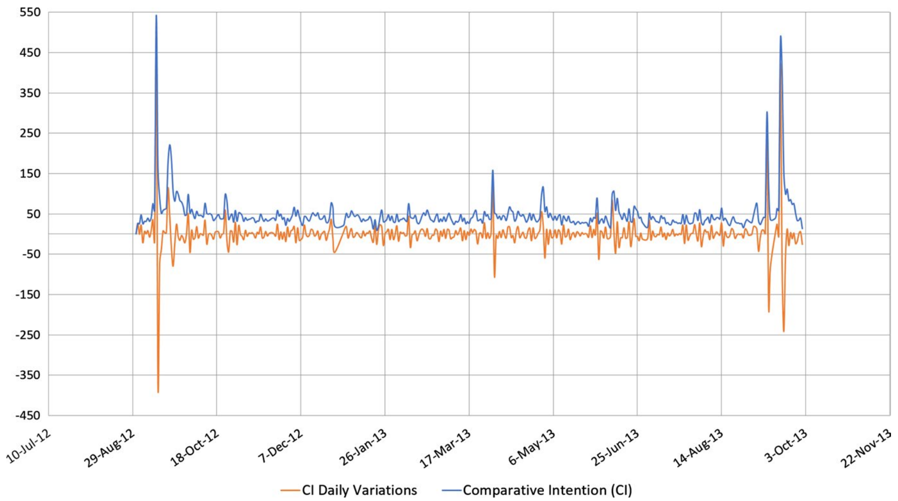

Fig. 14 Comparative intention (CI) and daily variations (change)

intention not to purchase the PoI, which bypasses the neutral and negative feelings. That reasonable conclusion is verified through the data.

Objectivity or subjectivity has similarities in both intentions (Purchase and no Purchase). We notice in Figure 13 that objectivity prevails; however, these conclusions are reverse to the hypothesis that most people express their personal opinions through social platforms as statements.

5.2 Comparative intention results

The fundamental approaches in the field of goods with a rapidly changing lifecycle are mentioned by Sharma (2013). By studying these approaches followed

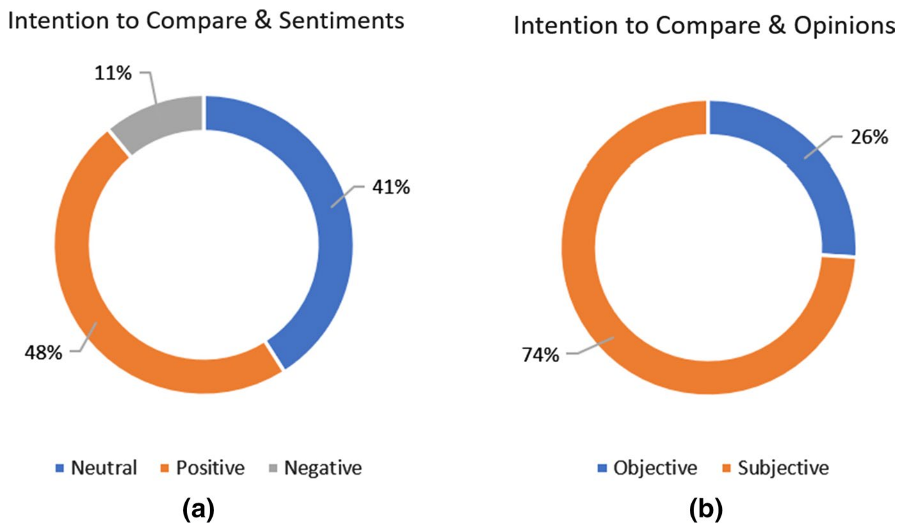

Fig. 15 Comparative intention and sentiment (**a**), comparative intention and opinion (**b**)

by companies, it is obvious how the insights from the comparison intent tweets can be significant.

Concerning the contribution to advertising, the company needs to support the brand's differences from other competitors during the product maturity phase. Knowing the features that each competitor opposes according to the public opinion, the company can make targeted campaigns.

In Fig. 14, we identify the changes in the users' comparison intentions. Obviously, during the maturity phase of the PoI, there are constant variations in the number of tweets. Therefore, the company should study these spikes in intention to compare, analyze the tweets returned by the model, and make decisions.

Nevertheless, also for the product itself, during its development, the company can identify what feature needs improvement by studying the comparison of products made by users based on specific characteristics. In particular, a small analysis of the data proved that users had an issue with their chargers, but mainly with the battery life, and that was the main reason they wanted Nokia phones over the iPhone 5. Regarding the sentiment polarization and the opinion, we present the similarity in Fig. 15.

The emotion that describes this intention to compare based on a product feature is mainly positive or neutral. At the same time, subjectivity dominates with almost 75% as users submit their personal opinion when they compare two products. Definitely, some cases can be objective, for example, when the comparison is stated based on a feature.

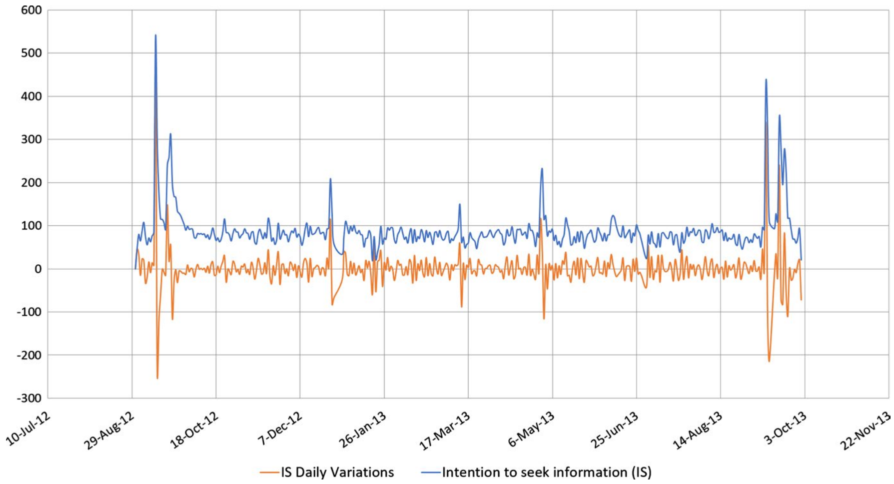

Fig. 16 Intention to seek information during the product life-cycle

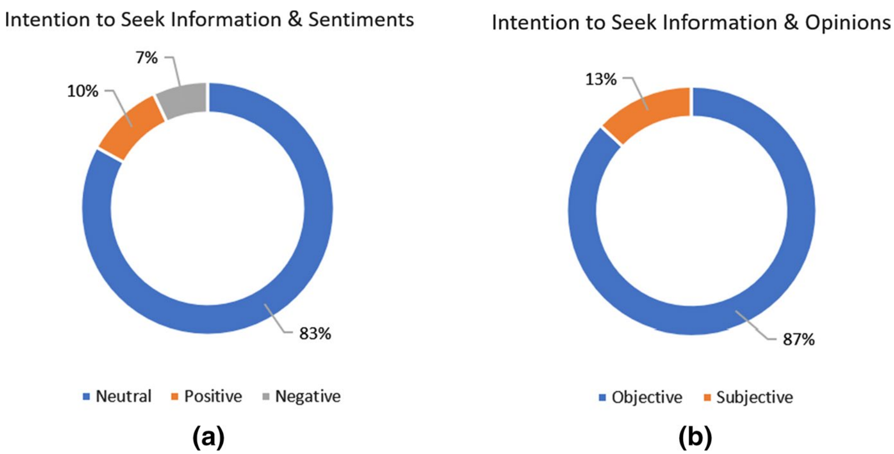

Fig. 17 Intention to seek information and sentiment (**a**), comparative intention and opinion (**b**)

5.3 Information seeking intention results

Users' intention to ask for information regarding the PoI allows the company to measure the public interest. Nowadays, it is imperative to study the specific intention and the answers given to the user. The reason is that users are influenced by the answers they receive to questions about goods, when these they come from people they trust.

Observing Fig. 16, it seems that users' intention to seek information remains almost constant. Some points worth mentioning are the phase of the introduction of the iPhone 5 and the new models to the market. In addition, a general consumer interest seems to exist on specific dates, such as Christmas.

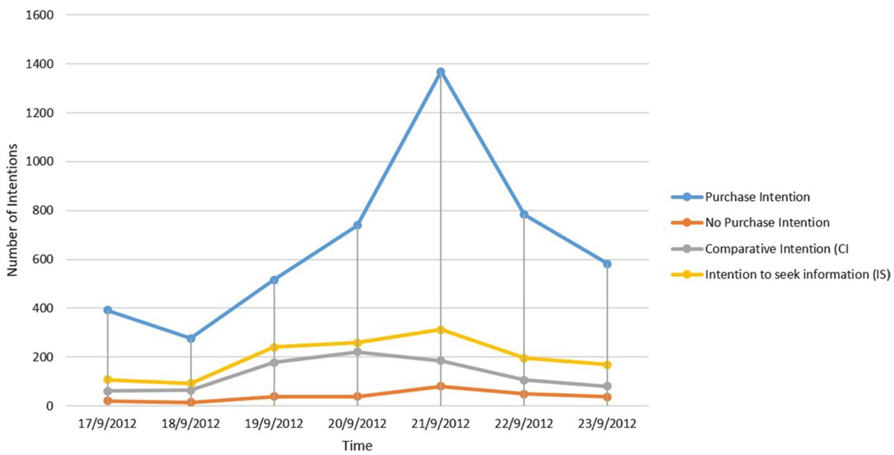

Fig. 18 Intent extraction in Samsung's aggressive campaign

Sentiment and opinion analyses of the tweets in Fig. 17 showed that the set of questions does not contain any emotion. That is in line with reality, as it is rare that one conveys emotion through a question. Also, when people want to learn something, they usually ask concisely and clearly. These features are found in the data, as there is no particular sentiment polarization, and objectivity prevails.

5.4 Tweet examples and case studies

During this research, we worked with an unlabeled dataset; this complicates the evaluation process. For this reason, we initially tried to study examples of tweets indicating user intention. Another way to evaluate system's quality is from the end-user's perspective, checking if the system can detect any information helpful to the business. Evaluating tweets by category is relatively easy, as it is enough to export a few examples. By doing so, we observed, as expected, that there were cases in which the system worked satisfactorily and others in which correct intention detection was difficult.

While observing the intentions' diagram study, part of which is shown in Fig. 18, we observed an unexplained increase in the number of intentions on a specific date. Further investigation showed that, on this date, the company did not make any market move, e.g. announcements or introduction of a new model or feature. Therefore, the proposed system had detected an external incident that affected the studied PoI and, as a result, the business as a whole. Specifically, on 9/20/2012, Samsung released its first TV commercial (aired on NBC), where it once again mocks Apple fanboys queuing outside Apple Stores, saying that "the next best thing is already here and is the Samsung Galaxy S III". However, according to reports at the time, it did not have positive results for Samsung, as users supported the iPhone 5. Table 11 presents example tweets' as classified in intention categories from the proposed framework.

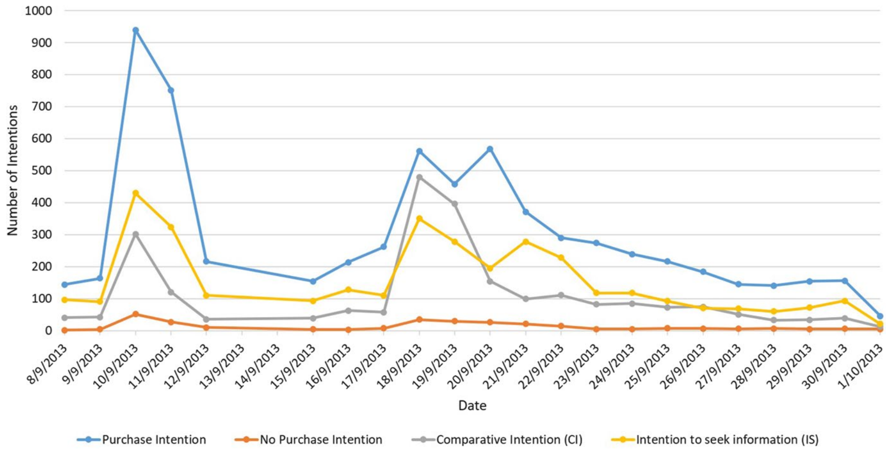

Fig. 19 Mining of intentions during the announcement of Apple iPhone 5c, 5s models (10-September-2013)

Table 11 Case study's example Tweets

Intention category	Tweet
Comparative intention	Samsung S3 + iPhone are more or less the same now! But the S3 got alot more things i prefer! E.g. BlueTooth!! Battery Life + easier 2 use
Purchase intention	I already got a nexus. i want an rbz. but i wanna buy the iphone 5 or a samsun
Information seeking intention	Which one should I choose iphone or S3?

Besides the identification of external incidents, the system was also triggered by a company's internal decision. Such an event occurred on September 18–20, 2013, there was an increase in all intentions, but the comparative intention was critical for the company, as shown in Fig. 19.

6 Conclusions and future work

6.1 Conclusions

This paper studied the processes for extracting user intentions from microblogs and aimed to collect and analyze user-generated content publicly available and accessible through Twitter. While studying the problem, many limitations and opportunities that could have a BI tool for a business were identified. There are two types of such systems; those that rely on processing vast volumes of data and those with high accuracy. A few hundred thousand tweets were studied, starting with a set of data consisting of billions of them. So the proposed framework belongs to the first category, as it processes a massive amount of data coming from Twitter.

The proposed framework used automated procedures to extract users' intentions and detect sentiment polarization and opinions from tweets related to the PoI. This framework was implemented based on iPhone 5, but it allows it to analyze any tweet related to a product with appropriate input modifications. As a result, we have collected, analyzed, and proved the importance of these data for a business. In addition, we have proposed how this unsupervised method can be transformed into a supervised method. Specifically, the framework can be employed unsupervised during a product's life-cycle, e.g., iPhone 5, and create a small portion of data that can be easily labelled. Following that, these data can be used to train a multi-label classification model that can automatically classify tweets when a new PoI enters the market.

The following subsections present some managerial implications about the proposed framework, how to digitize an SME, and limitations and opportunities for the future scope of the framework.

6.2 Managerial implications and main concerns

Approaching the findings mentioned above from a managerial perspective, managers should pay attention to online reviews of Twitter or various social media websites in different periods and PoIs'. This way, managers can gain a significant insight into their customers' intentions and information regarding the quality of their products, which will improve the product's attributes, as Liu et al. (2017a) also state. Managers should also consider that they can utilize sentiment and positive online textual reviews from customers to promote the benefits of their products and services, and they can use the positive online textual reviews to promote themselves as successful marketing paradigms as Xu et al. (2017) and Chatterjee (2019) also argue.

Further, the results from our framework can also help managers better provide customers with more information about the attributes of the PoI and identify different needs of customers depending on demographics. In this way, various marketing campaigns for specific groups will increase their sales. Managers should also be concerned about the volume of data and how they process them. The appropriate sampling of the data set and the suitable Natural Language Processing (NLP) methods will extract accurate customer satisfaction and dissatisfaction about a specific PoI.

In conclusion, this specific framework and its results provide business managers and SMEs with information about microblogs' user intentions. Business external or internal decisions impact on the public's opinion, intentions, and sentiment, as they are expressed on social networks. There is a feedback system between businesses and users from which valuable information can be extracted.

In Sect. 3, the main concerns are described and focus is placed on the study of user intention on social networking platforms and extraction of valuable information concerning a product of the proposed research. These concerns about the proposed study were answered by: collecting data from the product's life cycle of interest, the iPhone 5. The dataset contains tweets for each day of the iPhone 5's life cycle, a total of 365 days.

Then, the model was based on linguistic rules, machine learning, and the above combination. At the same time, a total of six different methodologies were implemented to extract intent, sentiment and type of opinion from a tweet based on existing literature. Insights regarding the changes in the system's intentions concerning the company's sales showed a direct connection between them. Finally, presenting a real example of introducing a competing product in the market helped assess the problem.

6.3 SMEs digitalization

Digitalization is becoming more and more necessary to businesses of all sizes and industries. The accompanying digital transformation of their respective business models has impacted small and medium-sized enterprises (SMEs), which have played an essential role in a country's economy (Becker and Schmid 2020). For SMEs to remain competitive, this requires rethinking their strategic direction and digital strategy and adapting them to the new underlying situational realities they face today. As a result, it can be anticipated that SME's digital strategy must include BI solutions like the proposed framework.

The following steps highlight the implementation of the proposed approach in details. In the first step, unsupervised mining can occur throughout the life of the PoI, providing a roadmap for enhancing service quality and company performance and more effectively targeting the market by implementing appropriate market strategies. Following, intents can be extracted and decision-making aided through quantitative and qualitative evaluations of the service components' impact on customer outcomes via a tool, for example (McKinney et al. 2010).

The next stage in a business is to use a subset of the data to train a weak-supervision tool, such as the one proposed by Ratner et al. (2020). This task can be accomplished by assigning an employee to identify/annotate a subset of extracted intentions and capable of grading a single piece of data. Following that, use a custom function to create an initial graded data set, and train supervised models using the labelled data.

Finally, businesses can use critical indicators to calibrate short and long-term outcomes using, at the same time, the real-time unsupervised methodology. In addition, the supervised model is trained to aid in future decision-making that will enable them to monitor consumer outcomes and be used as a BI tool.

6.4 Limitations and opportunities for future scope

First, the volume of data required to investigate the public intentions throughout the product's life cycle was tremendous, and data from Twitter for an entire year was required. Additionally, extracting the relevant tweets was time-consuming, as 1,600,000,000 tweets had to be analyzed serially.

Additionally, a significant challenge was remove the noise and extract intentions. Finally, the data rating process was not conducted in the research context because the total amount of data was too large, and the cost of rating it was prohibitively high. Additionally, there was a dearth of data knowledge. Without knowing the data set under study, appropriate questions for the raters cannot be formulated.

The present work's future directions begin with an adequate selection of the data set to build a graded set. Additionally, the dataset should be used to refine this automatic model. Furthermore, a comparative study of the automated detection procedure to the machine learning model could produce meaningful quantitative results. The framework's applicability in the data lifecycle of the following Apple products is another direction. Finally, Deep Learning Application is a discipline redefining data analysis and mining in the modern era.

References

Alfonseca E, Filippova K, Delort J, et al. (2012) Pattern learning for relation extraction with a hierarchical topic model. In: The 50th annual meeting of the association for computational linguistics, proceedings of the conference, July 8–14, 2012, Jeju Island, Korea-Vol 2: Short Papers. The Association for Computer Linguistics, pp 54–59. https://aclanthology.org/P12-2011/

Bagozzi RP (2010) Consumer intentions. Wiley, New York. https://doi.org/10.1002/9781444316568.wiem03057

Becker W, Schmid O (2020) The right digital strategy for your business: an empirical analysis of the design and implementation of digital strategies in smes and lses. Bus Res 13(3):985–1005

Blythe M, Cairns PA (2009) Critical methods and user generated content: the iphone on youtube. In: Jr. DRO, Arthur RB, Hinckley K, et al (eds) Proceedings of the 27th international conference on human factors in computing systems, CHI 2009, Boston, MA, USA, April 4–9, 2009. ACM, pp 1467–1476, https://doi.org/10.1145/1518701.1518923

Carlos CS, Yalamanchi M (2012) Intention analysis for sales, marketing and customer service. In: Kay M, Boitet C (eds) COLING 2012, 24th international conference on computational linguistics, proceedings of the conference: demonstration papers, 8–15 December 2012, Mumbai, India. Indian Institute of Technology Bombay, pp 33–40. https://www.aclweb.org/anthology/C12-3005/

Chatterjee S (2019) Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. Decis Support Syst 119:14–22. https://doi.org/10.1016/j.dss.2019.02.008

Chatterjee S, Krystyanczuk M (2017) Python social media analytics. Packt Publishing Ltd, London

Chen Z, Liu B, Hsu M, et al (2013) Identifying intention posts in discussion forums. In: Vanderwende L, III HD, Kirchhoff K (eds) Human language technologies: conference of the North American chapter of the association of computational linguistics, proceedings, June 9–14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA. The Association for Computational Linguistics, pp 1041–1050. https://www.aclweb.org/anthology/N13-1124/

Choi J, Yoon J, Chung J et al (2020) Social media analytics and business intelligence research: a systematic review. Inf Process Manag 57(6):102279. https://doi.org/10.1016/j.ipm.2020.102279

Ding X, Liu T, Duan J, et al (2015) Mining user consumption intention from social media using domain adaptive convolutional neural network. In: Bonet B, Koenig S (eds) Proceedings of the twenty-ninth AAAI conference on artificial intelligence, January 25–30, 2015, Austin, Texas, USA. AAAI Press, pp 2389–2395. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9748

Eckerson WW (2010) Performance dashboards: measuring, monitoring, and managing your business. Wiley, London

Faulds DJ, Mangold WG, Raju P et al (2018) The mobile shopping revolution: redefining the consumer decision process. Bus Horiz 61(2):323–338. https://doi.org/10.1016/j.bushor.2017.11.012

Felt M (2016) Social media and the social sciences: How researchers employ big data analytics. Big Data Soc 3(1):2053951716645,828. https://doi.org/10.1177/2053951716645828

Gao W, Sebastiani F (2016) From classification to quantification in tweet sentiment analysis. Soc Netw Anal Min 6(1):1–22

Giachanou A, Crestani F (2016) Like it or not: a survey of twitter sentiment analysis methods. ACM Comput Surv 49(2):28:1-28:41. https://doi.org/10.1145/2938640

Gupta V, Varshney D, Jhamtani H, et al (2014) Identifying purchase intent from social posts. In: Adar E, Resnick P, Choudhury MD, et al (eds) Proceedings of the eighth international conference on weblogs and social media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1–4, 2014. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8037

Guzman E, Maalej W (2014) How do users like this feature? A fine grained sentiment analysis of app reviews. In: Gorschek T, Lutz RR (eds) IEEE 22nd international requirements engineering conference, RE 2014, Karlskrona, Sweden, August 25–29, 2014. IEEE Computer Society, pp 153–162, 10.1109/RE.2014.6912257,

Hamroun M, Gouider MS, Said LB (2016) Large scale microblogging intentions analysis with pattern based approach. In: Howlett RJ, Jain LC, Gabrys B, et al (eds) Knowledge-based and intelligent information and engineering systems: proceedings of the 20th international conference KES-2016, York, UK, 5–7 September 2016, Procedia computer science, vol 96. Elsevier, pp 1249–1257. https://doi.org/10.1016/j.procs.2016.08.169

He W, Wu H, Yan G et al (2015) A novel social media competitive analytics framework with sentiment benchmarks. Inf Manag 52(7):801–812. https://doi.org/10.1016/j.im.2015.04.006

Hollerit B, Kröll M, Strohmaier M (2013) Towards linking buyers and sellers: detecting commercial intent on twitter. In: Carr L, Laender AHF, Lóscio BF, et al (eds) 22nd international world wide web conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013, Companion Volume. International World Wide Web Conferences Steering Committee / ACM, pp 629–632. https://doi.org/10.1145/2487788.2488009

Hung C, Lin H (2013) Using objective words in sentiwordnet to improve word-of-mouth sentiment classification. IEEE Intell Syst 28(2):47–54. https://doi.org/10.1109/MIS.2013.1

Hutto CJ, Gilbert E (2014) VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Adar E, Resnick P, Choudhury MD, et al (eds) Proceedings of the eighth international conference on weblogs and social media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1–4, 2014. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109

Jhamtani H, Chhaya N, Karwa S, et al (2015) Identifying suggestions for improvement of product features from online product reviews. In: Liu T, Scollon CN, Zhu W (eds) Social informatics-7th international conference, SocInfo 2015, Beijing, China, December 9–12, 2015, Proceedings, Lecture Notes in Computer Science, vol 9471. Springer, pp 112–119. https://doi.org/10.1007/978-3-319-27433-1_8

Jindal N, Liu B (2006) Identifying comparative sentences in text documents. In: Efthimiadis EN, Dumais ST, Hawking D, et al (eds) SIGIR 2006: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval, Seattle, Washington, USA, August 6–11, 2006. ACM, pp 244–251. https://doi.org/10.1145/1148170.1148215

Kalamatianos G, Symeonidis S, Mallis D et al (2018) Towards the creation of an emotion lexicon for microblogging. J Syst Inf Technol 20(2):130–151. https://doi.org/10.1108/JSIT-06-2017-0040

Kim Y, Dwivedi R, Zhang J et al (2016) Competitive intelligence in social media twitter: iphone 6 vs. galaxy S5. Online Inf Rev 40(1):42–61. https://doi.org/10.1108/OIR-03-2015-0068

Kumar N, Nagalla R, Marwah T, et al (2019) Sentiment dynamics in social media news channels. CoRR abs/1908.08147. arXiv:1908.08147

Kurnia PF, Suharjito (2018) Business intelligence model to analyze social media information. Procedia Comput Sci 135:5–14. https://doi.org/10.1016/j.procs.2018.08.144 (The 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018) : Empowering Smart Technology in Digital Era for a Bette Life)

Ladhari R, Michaud M (2015) ewom effects on hotel booking intentions, attitudes, trust, and website perceptions. Int J Hosp Manag 46:36–45. https://doi.org/10.1016/j.ijhm.2015.01.010

Li N, Wu DD (2010) Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decis Support Syst 48(2):354–368. https://doi.org/10.1016/j.dss.2009.09.003

Liu B (2010) Sentiment analysis and subjectivity. In: Indurkhya N, Damerau FJ (eds) Handbook of natural language processing, 2nd edn. Chapman and Hall/CRC, New York, pp 627–666. https://doi.org/10.1201/9781420085938-c26

Liu X, Burns AC, Hou Y (2017) An investigation of brand-related user-generated content on twitter. J Advert 46(2):236–247. https://doi.org/10.1080/00913367.2017.1297273

Liu Y, Bi J, Fan Z (2017) Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. Inf Fusion 36:149–161. https://doi.org/10.1016/j.inffus.2016.11.012

Lumezanu C, Feamster N (2012) Observing common spam in twitter and email. In: Proceedings of the 2012 internet measurement conference. Association for Computing Machinery, New York, NY, USA, IMC '12, pp 461–466. https://doi.org/10.1145/2398776.2398824

Manning CD, Surdeanu M, Bauer J, et al (2014) The stanford corenlp natural language processing toolkit. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, ACL 2014, June 22–27, 2014, Baltimore, MD, USA, System Demonstrations. The Association for Computer Linguistics, pp 55–60, https://doi.org/10.3115/v1/p14-5010

McKinney W, et al (2010) Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, Austin, TX, pp 51–56

Montoyo A, Martínez-Barco P, Balahur A (2012) Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. Decis Support Syst 53(4):675–679. https://doi.org/10.1016/j.dss.2012.05.022

Morris MR, Teevan J, Panovich K (2010) A comparison of information seeking using search engines and social networks. In: Cohen WW, Gosling S (eds) Proceedings of the fourth international conference on weblogs and social media, ICWSM 2010, Washington, DC, USA, May 23–26, 2010. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1518

Nashaat M, Ghosh A, Miller J, et al (2018) Hybridization of active learning and data programming for labeling large industrial datasets. In: IEEE international conference on big data (IEEE BigData 2018), Seattle, WA, USA, December 10–13, 2018. IEEE, pp 46–55, https://doi.org/10.1109/BigData.2018.8622459

Negash S, Gray P (2003) Business intelligence. In: 9th Americas conference on information systems, AMCIS 2003, Tampa, FL, USA, August 4–6, 2003. Association for Information Systems, p 423, http://aisel.aisnet.org/amcis2003/423

Paul SA, Hong L, Chi EH (2011) Is twitter a good place for asking questions? A characterization study. In: Adamic LA, Baeza-Yates R, Counts S (eds) Proceedings of the fifth international conference on weblogs and social media, Barcelona, Catalonia, Spain, July 17–21, 2011. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2813

Poecze F, Ebster C, Strauss C (2018) Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. In: Shakshuki EM, Yasar A (eds) The 9th international conference on ambient systems, networks and technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT 2018) / Affiliated Workshops, May 8–11, 2018, Porto, Portugal, Procedia Computer Science, vol 130. Elsevier, pp 660–666. https://doi.org/10.1016/j.procs.2018.04.117

Purohit H, Dong G, Shalin VL, et al (2015) Intent classification of short-text on social media. In: 2015 IEEE international conference on Smart City/SocialCom/SustainCom 2015, Chengdu, China, December 19–21, 2015. IEEE Computer Society, pp 222–228. https://doi.org/10.1109/SmartCity.2015.75

Ramanand J, Bhavsar K, Pedanekar N (2010) Wishful thinking-finding suggestions and 'buy' wishes from product reviews. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. Association for Computational Linguistics, Los Angeles, CA, pp 54–61. https://www.aclweb.org/anthology/W10-0207

Ratner A, Bach SH, Ehrenberg HR et al (2020) Snorkel: rapid training data creation with weak supervision. VLDB J 29(2–3):709–730. https://doi.org/10.1007/s00778-019-00552-1

Reeves M, Deimler MS (2009) Strategies for winning in the current and post-recession environment. Strategy Leadersh 2:1158

Rui H, Liu Y, Whinston AB (2013) Whose and what chatter matters? the effect of tweets on movie sales. Decis Support Syst 55(4):863–870. https://doi.org/10.1016/j.dss.2012.12.022

Sahoo SR, Gupta BB (2019) Hybrid approach for detection of malicious profiles in twitter. Comput Electr Eng 76:65–81. https://doi.org/10.1016/j.compeleceng.2019.03.003

Saif H, Fernández M, He Y, et al (2013) Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold. In: Battaglino C, Bosco C, Cambria E, et al (eds) Proceedings of the first international workshop on emotion and sentiment in social and expressive media: approaches and perspectives from AI (ESSEM 2013) A workshop of the XIII International Conference of the Italian Association for Artificial Intelligence (AI*IA 2013), Turin, Italy, December 3, 2013, CEUR Workshop Proceedings, vol 1096. CEUR-WS.org, pp 9–21. http://ceur-ws.org/Vol-1096/paper1.pdf

Salehan M, Kim DJ (2016) Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics. Decis Support Syst 81:30–40. https://doi.org/10.1016/j.dss.2015.10.006

Sharma N (2013) Marketing strategy on different stages plc and its marketing implications on fmcg products. Int J Mark Financ Serv Manag Res 2(3):121–136

Shukri SE, Yaghi RI, Aljarah I, et al (2015) Twitter sentiment analysis: A case study in the automotive industry. In: 2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), pp 1–5. https://doi.org/10.1109/AEECT.2015.7360594

Smith AN, Fischer E, Yongjian C (2012) How does brand-related user-generated content differ across Youtube, Facebook, and Twitter? J Interact Mark 26(2):102–113. https://doi.org/10.1016/j.intmar.2012.01.002

Stolcke A, Ries K, Coccaro N, et al (2000) Dialogue act modeling for automatic tagging and recognition of conversational speech. CoRR cs.CL/0006023. https://arxiv.org/abs/cs/0006023

Sun X, Zhang C, Li G et al (2018) Detecting users' anomalous emotion using social media for business intelligence. J Comput Sci 25:193–200. https://doi.org/10.1016/j.jocs.2017.05.029

Symeonidis S, Effrosynidis D, Arampatzis A (2018) A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. Expert Syst Appl 110:298–310. https://doi.org/10.1016/j.eswa.2018.06.022

Tsou M, Zhang H, Jung C (2017) Identifying data noises, user biases, and system errors in geo-tagged twitter messages (tweets). CoRR abs/1712.02433. arXiv:1712.02433

Tutunea MF, Rus RV (2012) Business intelligence solutions for sme's. Procedia Econ Finance 3:865–870. https://doi.org/10.1016/S2212-5671(12)00242-0 (International Conference Emerging Markets Queries in Finance and Business, Petru Maior University of Tîrgu-Mures, ROMANIA, October 24th - 27th, 2012)

Varma P, Ré C (2018) Snuba: automating weak supervision to label training data. Proc VLDB Endow 12(3):223–236. https://doi.org/10.14778/3291264.3291268

Vidya NA, Fanany MI, Budi I (2015) Twitter sentiment to analyze net brand reputation of mobile phone providers. Procedia Comput Sci 72:519–526. https://doi.org/10.1016/j.procs.2015.12.159 (The Third Information Systems International Conference 2015)

Wang L, Yan J, Lin J et al (2017) Let the users tell the truth: self-disclosure intention and self-disclosure honesty in mobile social networking. Int J Inf Manag 37(1):1428–1440. https://doi.org/10.1016/j.ijinfomgt.2016.10.006

Watson HJ, Wixom BH (2007) The current state of business intelligence. Computer 40(9):96–99. https://doi.org/10.1109/MC.2007.331

Wiecek-Janka E, Papierz M, Kornecka M, et al (2017) Apple products: A discussion of the product life cycle. In: 4th international conference on management science and management innovation, pp 159–164

Wu T, Khan FM, Fisher TA, et al (2005) Posting act tagging using transformation-based learning. In: Lin TY, Ohsuga S, Liau C, et al (eds) Foundations of data mining and knowledge discovery, studies in computational intelligence, vol 6. Springer, pp 319–331. https://doi.org/10.1007/11498186_18

Wyrwoll C (2014) Social media-fundamentals, models, and ranking of user-generated content. Springer, Berlin. https://doi.org/10.1007/978-3-658-06984-1

Xu K, Liao SS, Li J et al (2011) Mining comparative opinions from customer reviews for competitive intelligence. Decis Support Syst 50(4):743–754. https://doi.org/10.1016/j.dss.2010.08.021

Xu X, Wang X, Li Y et al (2017) Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors. Int J Inf Manag 37(6):673–683. https://doi.org/10.1016/j.ijinfomgt.2017.06.004

Yao Y, Sun A (2016) Mobile phone name extraction from internet forums: a semi-supervised approach. World Wide Web 19(5):783–805. https://doi.org/10.1007/s11280-015-0361-1

Zhao Z, Mei Q (2013) Questions about questions: an empirical analysis of information needs on twitter. In: Schwabe D, Almeida VAF, Glaser H, et al (eds) 22nd international world wide web conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013. International World Wide Web Conferences Steering Committee / ACM, pp 1545–1556. https://doi.org/10.1145/2488388.2488523

Zheng ZE, Fader PS, Padmanabhan B (2012) From business intelligence to competitive intelligence: inferring competitive measures using augmented site-centric data. Inf Syst Res 23(3–1):698–720. https://doi.org/10.1287/isre.1110.0385