Establishing an automated brand index based on opinion mining – analysis of printed and social media

Abstract

This article analyses the potential of using opinion mining based on big data to calculate a brand index to reflect brand image in the media. The study is realised as a combination of analysing previous literature and applying a media monitoring tool to analyse editorial publications and social media to gain brand-related media sentiment. The potential of opinion mining and the use of vast amounts of data is demonstrated. The results indicate that sentiment analysis based on big data has potential for automating the calculation of brand indices. It seems that big data can be used to compare brands and the nature of their media visibility. Marketing research and the analytics domain can benefit from big data and their related meaningful applications.

1 Introduction

Brands have traditionally been listed and given an order based on a certain set of metrics. Forbes most valuable brand index is an example of a well-known brand index (Forbes, 2019). The Forbes Global 2000 is another example that is based on a mix of four metrics: sales, profit, assets and market value (Murphy, 2015). Assessing brands from a variety of perspectives is essential as relationships exist between the brands and the firm-level brand outcomes, the market share and relative price (Chaudhuri and Holbrook, 2001). Traditional printed media and different brands have had a long-standing relationship. Simultaneously, a variety of brand communities exist on social media, and the community contents are seen to affect customer behaviour (Goh *et al*, 2013). A variety of brand-related communication takes place in the media. However, the current brand indices do not consider the media sentiment of traditional published media or social media. The research is lacking in its understanding of the significance of brand-related media sentiment.

The role of media–brand relationships is not new to research, but social media growth provides new opportunities (Hofacker and Belanche, 2016). Brand–consumer interactions driven by social media have been studied (Rohm *et al*, 2013). Social media is seen to have a role in customer engagement (Chan *et al.*, 2014; Chauhan and Pillai, 2013). Positive and negative comments on social media are recognised to have a guiding role for brand-related communication (de Vries *et al*, 2012). In relation to the roles of traditional and social media, Bruhn *et al*. (2012) found traditional media to have a stronger impact on brand awareness, whereas social media communications strongly influence brand image. Regardless of the interest in the relationship between media communication and brand-related considerations, previous studies have not attempted to harness brand-related communication to analyse and compare different brands on a larger scale.

Comparing brands, or analysing media communication in relation to any topic is possible by analysing media sentiment. Opinion mining is a way of analysing the media sentiment related to a particular topic, by mining the vast data which are publicly available (Ahmed *et al*, 2015). The manual analysis of media sentiment of brand-related content is no longer practical due to the sheer volume of data (Dhaoui *et al*, 2017; Wang *et al*, 2012), and automated analysis is necessary. The computational analysis of vast amounts of data has only recently become truly viable due to developments in information technology (Chen *et al*, 2012). In a marketing context, traditional marketing relies on analytics that relate to very small data sets and limited analytic capacity (Xu *et al*, 2016). Big data refers to vast, often unstructured data sets that have been impossible to analyse in the past. Much of the big data-related focus has been unclear, and much hype is also associated with the term (Weinberg *et al*, 2013). Currently, the potential of big data is underutilised in the marketing domain, and particular efforts with a clear focus are needed. Brand analysis by means of media sentiment of traditional and social media might have the potential for meaningful focus.

The purpose of this study is to explore the possibilities of creating an automated brand image based on a brand index. The study applies algorithm-based computational linguistics to mine global brand sentiment. The data collection performs advanced media analysis on a vast number of editorial and social media sources. The study attempts to show the potential of opinion mining for the above-mentioned purpose, providing industry sector-related examples to support the logic and carrying out a level of comparison to an existing brand index. A direct formula for calculating the sentiment-based brand index is not provided.

2 Brief Literature Review

It is suggested that brands can be valued using a variety of techniques. However, like any tool, each technique may be suitable for certain purposes only (Salinas and Amber, 2009). Most techniques are variations of three general approaches to brand valuation involving the following: 1) cost (the brand is valued on the basis of the historical cost of creation or cost of recreating a similar brand), 2) market (the brand value is referenced to open market values) and 3) income (based on the premise that future cash flows dictate brand value). Simultaneously, both traditional and social media communications seem to have an impact on brand equity, with traditional media having a stronger impact on brand awareness, and social media influencing brand image (Bruhn *et al*, 2012). Potential gains exist in relation to existing customers as well as for non-customers in terms of perceived brand image or the perceived brand value via the increasing social media involvement (Zauner *et al*, 2012). Nevertheless, predictive analytic models have been developed to link the value that customers represent to a firm to customer profitability and other factors (Echchakoui, 2018).

Digital commerce and the rise of digital analytics have helped to address consumers' insatiable desire for immediate gratification (Fulgoni, 2018). A variety of digital applications have been developed in marketing research and in the field of market structure analysis to deal with classic challenges such as pricing, campaign management and brand positioning. Also, new product development is rooted in the analysis of product

substitutes and complements market structure methods in analysis and visualisation by automatically eliciting product attributes and brands' relative positions from online customer reviews (Lee *et al*, 2011). Social tags, which are user-defined keywords associated with online content are also used to reflect consumers' perceptions of products and brands (Nam *et al*, 2017). Various tools can be used to support business-to-business brands, including the Internet and other interactive technologies (Michaelidou *et al*, 2011). By focusing on measuring, brand owners can achieve increased effectiveness in understanding and optimising the way in which they leverage their media presence on social media, too (Lipsman *et al*, 2012). For example, McCorcindale (2013) suggests that a large social media visibility is required for actions at a concrete level. There are indications that brand communities, established on social media, have positive effects on shared consciousness, shared rituals and traditions and obligations to society, all of which have positive effects on value creation. Such communities have the potential to enhance brand loyalty through brand use and impression management practices (Laroche *et al*, 2012).

Brand equity is defined as: *"The differential effect of brand knowledge on consumer response to the marketing of the brand"* (Keller, 1993). The definition suggests that consumer assessments concerning a product with a brand name should be compared to an unnamed product without a brand. Previous research describes the benefits of branding for business-to-business organisations well, but research on how organisations use social networking to achieve brand objectives remains limited (Michaelidou *et al*, 2011). The relation of retail brand awareness and perceived quality to retail brand equity and certain managerial values has been analysed (Jara and Cliquet, 2011). The importance of retail attributes for retail brand equity is seen to vary among different sectors, but a link between equity and consumers' loyalty is evident (Swoboda *et al*, 2016). Consumer-based brand equity can generally be classified as either direct or indirect (Christodoulides and De Chernatony, 2010). Brand value is seen to depend on several stakeholders that function as a network to support or work against the brand value (Jones, 2005). Online reviews are seen to have a significant impact on brand awareness, perceived value and organisational associations, which lead to purchase intentions (Chakraborty and Bhat, 2018).

Traditional brand listings usually indicate an estimate of brand value and visibility. Most of the current brand indices are based on different subjective estimation methods and questionnaires. Considering opinion mining and machine learning-based methods to measure brand image requires the clarification of some concepts. Brand image can be defined as*: "The current view of the customers about a brand. It can be a unique bundle of associations within the minds of customers. It is the customers' perception and the way a specific brand is positioned in the market"* (MSG, 2016). Therefore, brand image conveys emotional value and not just a mental image. Image has been defined as: *"The set of beliefs, ideas and impressions that a person holds of an object"* (Barich and Kotler, 1991). The definition of image is also seen as: *"Human construct imposed on an array of perceived attributes projected by an object, event, or person. A definition which summarises the view that an image is (1) a subjective, mental construct, (2) affects how things are perceived but also (3) is influenced by projected messages."* (Nimmo and Savage, 1976). Brand image seems to be impacted by firm-created social

media communication, while user-generated social media communication is seen to influence the hedonic brand image (Bruhn *et al*, 2012). User opinions can be determined by the means of sentiment analysis, which is increasing in social media (Chaudhary *et al*, 2018). Brand experience relates to sensations, feelings, cognitions and behavioural responses evoked by brand-related stimuli, made up of a brand's design and identity, packaging, communications and environment (Brakus *et al*, 2009). Previous studies have applied sentiment analysis to online reputation management (Olaleye *et al*, 2018), to improve energy sector-related decision-making (Nuortimo and Harkonen, 2019), and to analyse the media image of different technologies (Nuortimo *et al*., 2017) to give some examples. The related data-driven analytics and decision-making have also been discussed (Krishen and Petrescu, 2019).

3 Method: Opinion Mining via Automated Learning Machine-Based Media Analysis

This study explores the possibilities of creating an automated brand index by applying algorithm-based computational linguistics to mine global brand sentiment in printed and social media. The literature is also reviewed to reveal some of the previous brand-related discussion. This study, however, does not attempt to explain or take a position on the brand-related chains of reasoning. The intention is merely to demonstrate the opportunities that opinion mining may offer for analysing brands and reveal some related challenges.

Advanced media analysis is performed on a vast number of editorial and social media sources. Media sentiment is defined based on a large dataset with automatic algorithms. Sentiment analysis reveals people's opinions, comments, posts, reviews, evaluations and emotions, which are expressed about products, sales, services, individuals, organisations and such (Kumar *et al*, 2017). The scope of this paper is an exploratory method application, and the level is descriptive. Brand-related media sentiment that is based on big data classification and clustering, including both structured and unstructured data, reveals how the brands are perceived in the media.

It must be understood that, even now, the available computational capacity sets limitations for the possible analysis. For this reason, the potential of analysing brand-related media sentiment is explored in stages: 1) Industrial branch analysis is carried out to analyse the top 100 Finnish start-up companies. This branch analysis is intended as an example to enable understanding of the logic and potential of big data-based sentiment analysis. The start-up analysis shows that data are a prerequisite for the analysis. 2) Second industrial branch analysis is carried out to analyse global power equipment manufacturers. This industry sector-related example further shows the potential of analysing brand-related media sentiment. This type of analyses, limited to an industry sector, is possible with reasonable computational capacity. 3) The potential of sentiment analysis based on big data for brand index type of use is demonstrated by applying sentiment analysis to two companies on the Forbes list and discussing the analysis findings against the positioning of these two companies. The companies are selected so that they are not at the top of the Forbes list to ensure the possibility of working

with data sets that it is possible to deal with using the available computational capacity. 4) The analysis findings are synthesised to discuss potential logic for a sentiment-based brand index to reflect brand image. A direct formula for calculating a sentiment-based brand index is not provided. 5) Limited validation of the idea of a sentiment-based brand index is attempted by interviewing six people working in different fields. The idea is demonstrated for the interviewees before the interviews.

Sentiment analysis is not new; it was carried out manually before adequate computational possibilities existed. It has been applied to various fields including marketing research. The growth of social networks has increased the possibilities of this type of analysis, which has given rise to domain adaptation as a prominent area of research in sentiment analysis (Deshmukh *et al.*, 2017). Also, editorial news reports and articles can contain subjective components that reflect the necessary views, values and beliefs (Hillert *et al*, 2018). Sentiment analysis, word recognition and textual analysis have also been rapidly increasing in financial and economic research lately (Loughran and McDonald, 2016). This study analyses both social media and the more traditional news media.

The utilised opinion mining, based on learning machine-based analysis, is carried out by commercial M-adaptive software. Big data is utilised, both from open public sources (unstructured data) as well as from payable editorial sources (structured data). The sentiment of a document is based on different lexicons for different languages, where the algorithm calculates local sentiments of a document, and compares those to the search terms, bringing the analysis to aspect level. Then, the result is presented for the entire document to indicate positive, negative, neutral, mixed or unknown sentiments. By combining these sentiments and clustering the findings, an overall picture from machine-defined opinion can be observed. MS Excel software is utilised for clustering. The sentiments from editorial and social media sources are clustered separately to highlight the differences. The utilised software provides about eighty percent accuracy. It must be understood that machine learning techniques are not perfect as they have only reached a ninety percent accuracy level, with the average being just over seventy percent, depending on the algorithms used (Ball and Elworthy, 2014). Nevertheless, the accuracy of human analysis also has its limitations.

The analysed sources cover 3 million social media platforms and 100,000 news outlets in 71 languages and 236 regions. The scale of analysed sources varies from local discussion forums involving people living in small towns, to the Financial Times, a major influential newspaper. The inclusion of both social media and editorial content provides some extra insight into the brand image, which is not necessarily present in traditional studies. The identification of trends in the data becomes increasingly accurate with the increase in the number of analysed documents. The analysis is possible for anyone to repeat as a variety of open access software exists as well as commercial software tools (Kumar *et al.*, 2017). The limitation is typically the available computational capacity. If different software is used, it, as well as the extent of the available sources, might be a source of some variation. Some of the best-known software was tested previously to compare the payable and free software (Chen, 2018).

4 Brand-Related Media Sentiment Analysis

This section presents and discusses the results of brand-related media sentiment analysis for the following: 1) industrial branch analysis of the top 100 Finnish start-up companies, 2) industrial branch analysis of global power equipment manufacturers, 3) demonstrates the potential of sentiment analysis based on big data for brand index type use, 4) synthesises the analysis findings to discuss potential logic for a sentiment-based brand index to reflect brand image and 5) carries out limited validation of the idea of a sentiment-based brand index.

4.1 Industrial branch analysis – Top 100 Finnish start-up companies

Media sentiment analysis was first carried out on the top 100 start-up companies in Finland. The companies were obtained from the start-up 100 list (Start-up, 2016). The company names were used as keywords for the analysis. The amount of data points was so small that adequate data could not be obtained for meaningful analysis. This might be due to the companies being very unknown globally. Even the hottest start-ups had only a few hits per year, counting both social media and editorial media. These small amounts of data would be more meaningful if analysed manually. The analysis was not continued. This attempted analysis of the top 100 Finnish start-up companies indicates that media sentiment analysis on companies with limited media exposure is not feasible. The start-up analysis shows that the existence of data in the form of media hits is a prerequisite for the analysis.

4.2 Second industrial branch analysis – global power equipment manufacturers

The global power equipment manufacturers can be compared based on analysed media sentiment. The results of the second industrial group branch analysis are presented in Figures 1 and 2. The global power equipment manufacturers are used as examples that provide a suitable quantity of media hits with reasonable software back-end capacity. One challenge of crunching global big data is the server capacity, which may limit the data sets and selected companies. The analysis period in this case is one year.
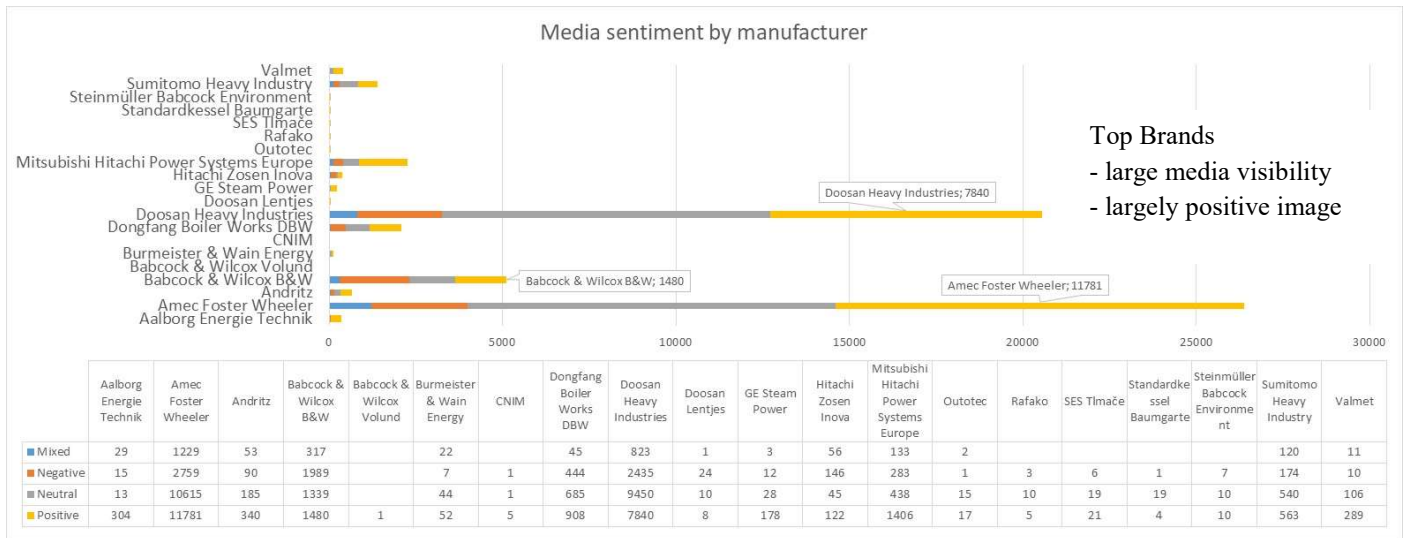
Figure 1. Media sentiment of power equipment manufacturers

	Aalborg Energie Technik	Amec Foster Wheeler	Andritz	Babcock & Wilcox B&W	Babcock & Wilcox Volund	Burmeister & Wain Energy	CNIM	Dongfang Boiler Works DBW	Doosan Heavy Industries	Doosan Lentjes	GE Steam Power	Hitachi Zosen Inova	Mitsubishi Hitachi Power Systems Europe	Outotec	Rafako	SES Tlmače	Standardkessel Baumgarte	Steinmüller Babcock Environment	Sumitomo Heavy Industry	Valmet
Mixed	29	1229	53	317		22		45	823	1	3	56	133	2					120	11
Negative	15	2759	90	1989		7	1	444	2435	24	12	146	283	1	3	6	1	7	174	10
Neutral	13	10615	185	1339		44	1	685	9450	10	28	45	438	15	10	19	19	10	540	106
Positive	304	11781	340	1480	1	52	5	908	7840	8	178	122	1406	17	5	21	4	10	563	289

Figure 1 shows the total number of media hits without making a distinction between the media sources. The total number of media hits for each company is the sum of positive, neutral, negative and mixed hits. The top three power equipment manufacturers by volume of media hits and the number of positive hits are Amec Foster Wheeler, Doosan Heavy Industries and Babcock & Wilcox B&W. If considering the positive hits only, noteworthy is that Mitsubishi Hitachi Power Systems Europe has nearly as many positive hits as Babcock & Wilcox B&W. Taking the accuracy of the media sentiment analysis, the question would be whether the difference fits within the margin of error. This illustration shows the differences in magnitudes of media hits between different companies. Amec Foster Wheeler and Doosan Heavy Industries have the highest numbers of neutral media hits. The media sentiment of this type of hit is neither positive nor negative, but merely neutral media visibility. Amec Foster Wheeler, Doosan Heavy Industries and Babcock & Wilcox B&W also obtain the highest number of negative media hits. The ratio of negative hits is relatively small against the total volume of media hits or the number of positive hits. In the case of Babcock & Wilcox B&W, the ratio of negative hits is the largest, relatively, compared to total and positive media hits. Should the companies be compared and indexed based on their media sentiment, a negative scoring might be necessary based on the number of negative media hits. The mixed hits are neither neutral, positive nor negative and cannot be classified into these categories. The number of mixed hits is the smallest, relatively, for each company. The volume of media hits and the nature of media sentiment in the industrial sector can be compared by this type of analysis.
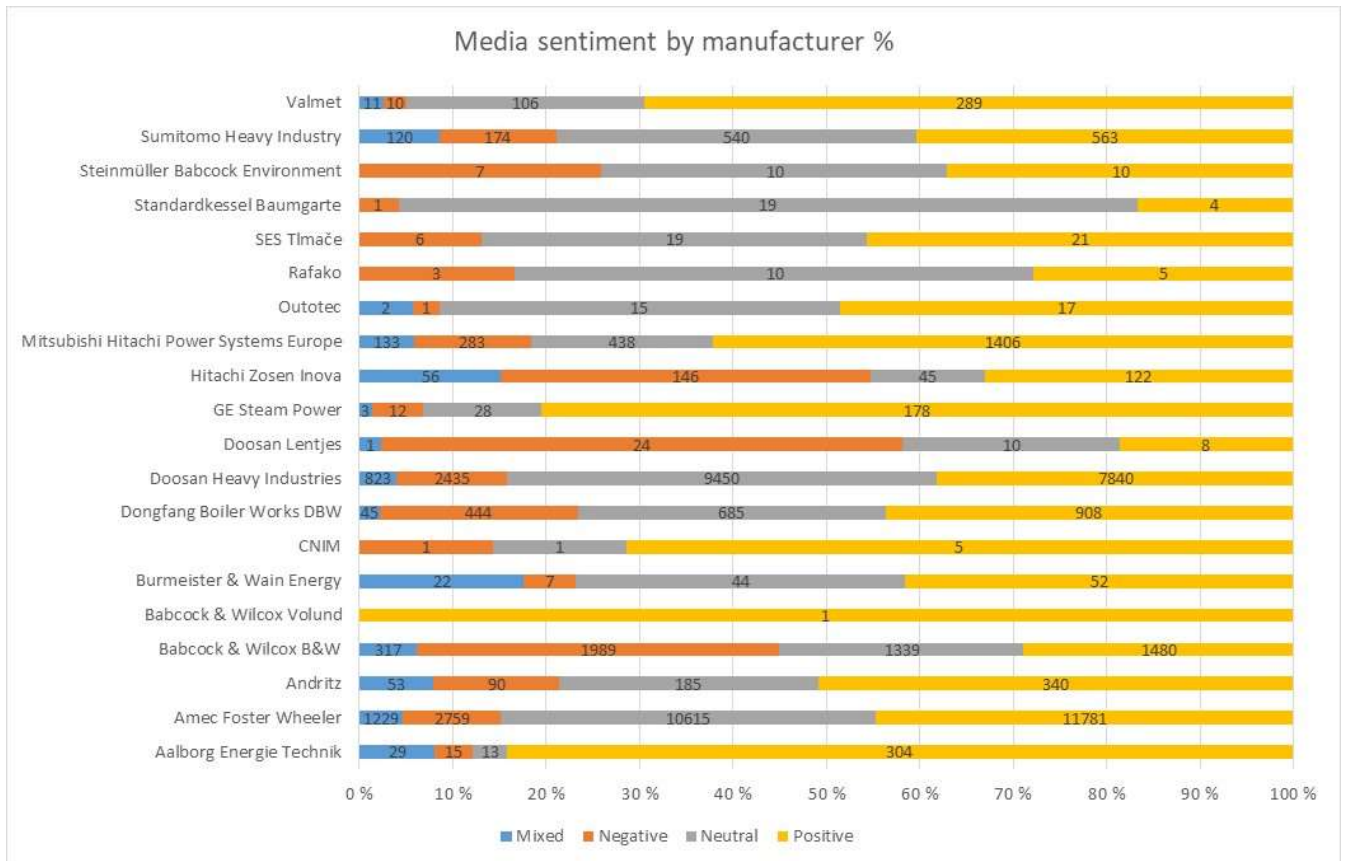
Figure 2. Percentual relative sentiment classification of media hits related to power equipment manufacturers

The same media sentiment can be presented in percentual shares of different types of media hits for each company (Figure 2). This type of presentation allows visual comparison of the shares of different types of hits between companies. The illustration also shows the total numbers of hits for each company. It can be seen that Aalborg Energie Technik has the largest share of positive hits in relation to its total number of media hits, compared to other companies. Babcock & Wilcox Volund has only one hit, a positive one, and should be ignored for this reason. GE Steam Power and CNIM have the next largest shares of positive hits compared to their total media hits. Valmet would be fourth if companies were to be compared in this manner. Doosan Lentjes has the largest share of negative hits compared to its total media hits. The nature of the media hits relates to the brand image according to some mechanisms.

The results indicate that brand visibility in the media can be analysed by media sentiment analysis. Media sentiment analysis provides information about the quantity of media hits and the nature of such hits. The underlying reasons for certain types of hits would, however, require further analysis. One of the challenges of analysis of this nature is that it does not directly reveal how much more influential negative sentiment is than positive sentiment. Media sentiment should be seen as directional in terms of the brand image. This industry sector-related example further shows the potential of analysing brand-related media sentiment but does not provide a direct formula for calculating a sentiment-based brand index.

4.3 Comparing sentiment analysis to the Forbes brand index

The media sentiment of two companies selected from the Forbes brand index (Forbes, 2019) are analysed to enable comparison of the results of the media analysis to the index. The companies Huawei and Adidas are intentionally selected so that they are known but are not at the top of the Forbes brand index, to ensure the sufficiency of the computational capacity. Also, the analysis period is limited to one month to ensure that an adequate number of hits are gained without excessive need for computational capacity.

Table 1 illustrates the position of the two companies on the Forbes list together with a comparison to opinion-mined data. The table includes the total number of media hits, the percentual share of negative hits against the total number of hits and the percentual share of negative hits in social media.

Table 1. Forbes brand index compared to opinion mining

Company	Forbes brand index		Data from opinion mining (one month)		
	Position on Forbes list	Brand value	Total media hits	Total hits: % negative	Social media hits: % negative
Huawei	88	$7.3 B	4959	15 %	29.87 %
Adidas	75	$7.9 B	4920	14.98 %	20.89 %

From Table 1, it is visible that Adidas has better status in the Forbes brand index than Huawei. However, Huawei has slightly more media hits in total, so based on the number of total hits, the Huawei brand image would be better for the analysed period. The percentual share of negative media hits in total is nearly identical for the two companies. The question is whether the difference fits within the margin of error. The clearest difference between the two companies lies in the percentual share of negative hits on social media. Adidas has received roughly 9% fewer negative hits on social media. Hence, it appears that the order of the companies, if judged based on media sentiment, could be set by using negative scoring based on negative media sentiment. The order of the two companies would likely be the same, with Adidas before Huawei. The data, however, do not reveal how the positioning would be among the other companies on the Forbes brand index.

Figure 3 presents the more detailed sentiment analysis for Huawei. Media sentiment is illustrated for the editorial publications and, separately, for different types of social media.
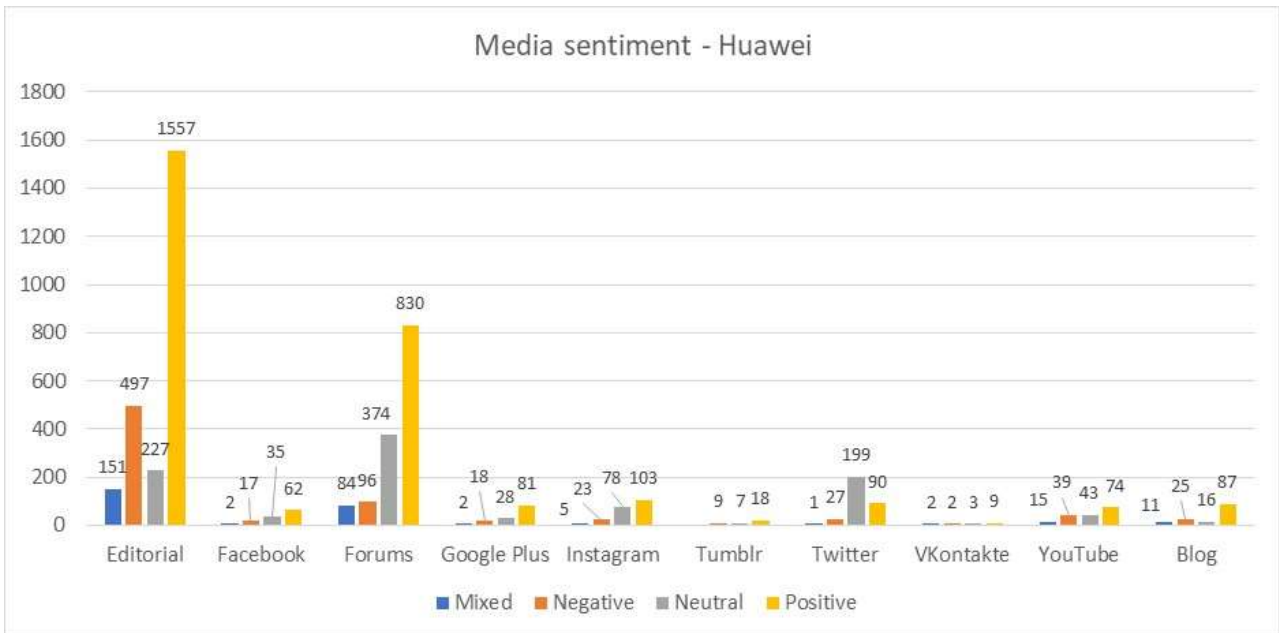
Figure 3. One-month global media visibility with sentiment classification – Huawei

Figure 3 shows the shares of positive, negative and neutral hits in different types of media for Huawei during the analysis period. The largest share of positive hits was in the editorial publications (1557 hits), which is slightly higher than the total number of those on social media, if all the social media are considered together (1354 hits). The total number of negative hits was also higher in the editorial media (497 hits), compared to the total negative hits on social media (256 hits).

Figure 4 presents a more detailed sentiment analysis for Adidas. Media sentiment is illustrated for the editorial publications, and, separately, for different types of social media.
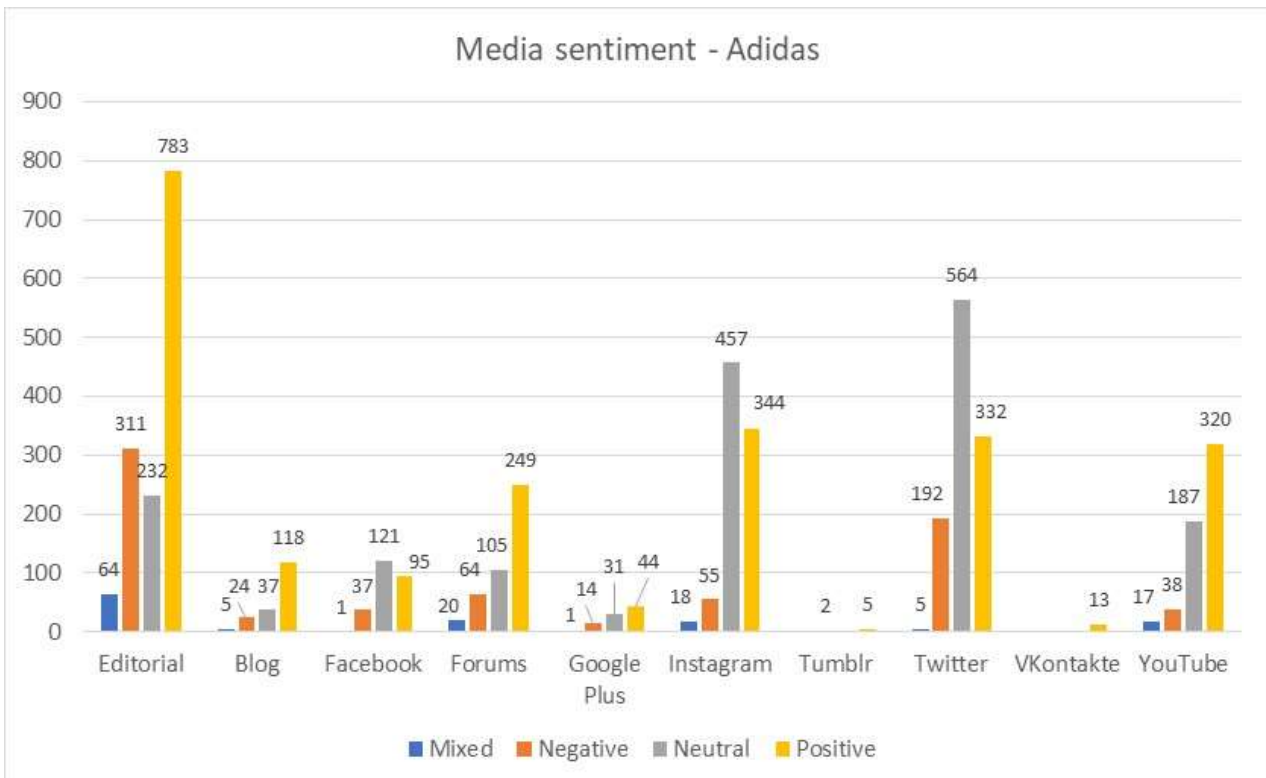
Figure 4. One-month global media visibility with sentiment classification – Adidas

Figure 4 shows the share of positive, negative and neutral hits for Adidas in different types of media during the analysis period. The largest share of positive hits was in the editorial publications (783 hits), which is nearly half the number of those on social media, if all the social media are considered together (1520 hits). The total number of negative hits was also fewer for Adidas in the editorial media (311 hits), compared to the total negative hits on social media (426 hits). With regard to social media, Twitter, Instagram and YouTube are emphasised in the case of Adidas.

Figures 3 and 4 enable a comparison of the media sentiments of Huawei and Adidas in the editorial media and on different social media during the analysis period. In the editorial media, Huawei obtained more positive attention than Adidas during the analysis period, but also received more relative negative attention. The figures also indicate that attention on social media is emphasised through different channels for these two companies. The analysis, however, does not reveal the underlying reasons for the differences and would require further study.

4.4 Analysis synthesis

Overall, the findings indicate that the media sentiment of different brands could be utilised to create an index based on brand image. The principles for scoring or a formula would need to be set to compare different brands. Considering the total number of media hits is one possibility, but, in particular, considering negative scoring based on negative hits might make sense. This type of brand image index can be automated based on computational algorithms and calculated for a variety of time frames, or can even be ongoing. The brand image-related index is possible to calculate separately for brands that relate to different industrial sectors or for top brands. The main limitation is the available computational capacity.

Table 3 proposes a possible index structure for brand images. The total number of media hits during the analysis period should be acknowledged, and the negative hits should cause negative scoring. A media sentiment-based brand index is possible to realise, for example, at monthly intervals or for a year at a time. Significant computational capacity is, however, required for extensive periods. A formula is not proposed for this type of brand index.

Table 3. Proposed brand image index structure

Brand	Media sentiment (One year)			Company rank (1–100)
	Total media hits	% negative	% negative on social media	
Brand 1				
Brand 2				
Brand 3				
Brand 4				
Brand n				

Table 3 implies that the overall brand visibility would be the primary rating; however, the rating would be influenced by both overall negative hits and negative social media hits. The top company would be the one with the largest global visibility, with the smallest relative amount of overall negative hits and the smallest relative number of negative social media hits.

4.5 Limited validation

Limited validation of the idea of a sentiment-based brand index was carried out by interviewing people working in different fields. The idea was demonstrated for the interviewees before the interviews. The interviewees represented the publishing sector, media consulting and a venture capital company investing in start-ups. The interviewees were asked what they thought about the idea in general, and whether they would visit a website with an automated media sentiment-based brand index. The interviewees were also asked whether they would be willing to buy additional information once knowing their own position on the index, and whether they would be interested in buying media consultancy to increase their position on the index.

Some of the interviewees found it initially challenging to understand the idea behind a media sentiment-based brand index. They saw some potential in the index but were also interested in additional benefits. Most of the interviewees would be interested in visiting a website with such an index, while others were slightly more cautious and hoped to understand further benefits. The interviewees would generally be willing to consider buying additional information after knowing their position on the index but would require actionable additional value and maybe some way to benefit their decision-making. The underlying message was that there must be a benefit to motivate them to spend money. A representative of the publishing company was hoping to see a level of human validation. The companies would consider additional consultancy should it be available, and should the index be reliable. Based on the limited validation, it seems that a reserved interest towards a media sentiment-based brand index exists. Creating such an index does provide new opportunities for analytics, but requires motivation and potentially a business model should this type of index be created with profit in mind.

5 Discussion

This study explores the possibilities of creating automated brand image-related indices based on computational algorithm-based opinion mining. The potential of mining global brand sentiment by means of media analysis is demonstrated. Media analysis performed on a vast number of editorial and social media sources enables the inclusion of data points from a vast amount of big data in a manner that has not been possible before due to deficient computational capacity. It appears that brand image-related indices are possible to create should someone have the motivation to do so. The automated media analysis would enable the inclusion of tens of thousands of companies from different industrial and commercial sectors.

The analysis indicates that computational media sentiment analysis is not feasible for companies with limited media exposure. The existence of data in the form of media hits is a prerequisite for the analysis. The results

also indicate that industrial branch analysis is possible and that it would be possible to index brands based on this type of analysis. However, the analysis does not directly reveal how much more influential negative sentiment is than positive sentiment. Hence, media sentiment should be seen as directional, and indexing would necessitate using a formula. This study does not provide a direct formula for calculating a sentiment-based brand index. The potential of sentiment analysis based on big data for brand index type use is also demonstrated by comparing sentiment analysis to the existing Forbes brand index, which is based on completely different metrics. The analysis findings are synthesised to discuss the potential logic for a media sentiment-based brand index. Limited validation is carried out to see whether there would be interest in this type of index.

A media sentiment-based index of brand image is a new application of opinion mining and provides a potentially meaningful big data application. The index would likely benefit large multinational brands due to global benchmarking towards industry standards. However, these companies usually have their own departments for marketing analytics. Full automation of the index might not be the best initial approach due to inaccuracies in sentiment classification, and the fact that companies that provide these types of services still rely on human analytics. Thus, a hybrid approach might be the best starting point. The index structure might benefit from indicating both volume and media sentiment. The inclusion of both editorial media and social media is recommended.

5.1 Scientific implications

The main scientific implication of this study involves demonstrating the possibilities of utilising big data-based media sentiment analysis on brand image analysis. This type of big data analysis provides possibilities for computational mining of global brand sentiment from traditional media and social media, and also the related brand analytics. This study is, hence, in line with that of Chaudhuri and Holbrook (2001) in understanding the importance of assessing brands from a variety of perspectives. This study also supports previous literature (Goh *et al*, 2013; Hofacker and Belanche, 2016; Rohm *et al*, 2013) by understanding that media and brands have a relationship and that brand communications may affect customer behaviour. Particularly, the role of social media in relation to brands may offer new opportunities. However, this study does not take a position on the underlying logic, but merely attempts to provide new ideas for the analytics. The same applies for the role of social media in customer engagement (Chauhan and Pillai, 2013; Chan *et al*, 2014). This study understands that certain logic may apply but does not take a position on or try to assume the logic. De Vries *et al* (2012) are given support by understanding the role of social media in brand communications, and a new contribution is provided by demonstrating the possibilities of utilising algorithm-based computational sentiment analysis. Particularly, Bruhn *et al* (2012) are provided with support by highlighting that both traditional and social media have a role to play in terms of brands. New analytical perspectives are provided in this study for considering both. The manner in which the type of media affects the relationship with brands is, however, not analysed.

Opinion mining and media sentiment analysis, in general (Ahmed *et al*, 2015; Chaudhary *et al*, 2018; Nuortimo *et al*., 2017; Nuortimo and Harkonen, 2019; Olaleye *et al*, 2018), are provided with support by introducing a new application. This study is in line with Dhaoui *et al* (2017) and Wang *et al* (2012) in understanding how manual analysis is not viable. However, this study indicates that manual analysis may still have a role. This study also supports research on data-driven analytics and decision-making (Krishen and Petrescu, 2019) by providing one example of an approach to data-driven brand analytics. Marketing analytics, previously relying on limited analytic capacity (Xu *et al*, 2016) are provided with alternatives and contributions, through a realistic example of utilising large data sets for brand related analytics. Yet, this study understands how the big data-related focus, in many cases, has been unclear (Weinberg *et al*, 2013) and provides its stake for clarifying the focus in the area of analysing brands. This study contributes by introducing new means of analytics to observe brands but also understands that any technique for analysing brands is suitable only for certain purposes and, hence, supports Salinas and Amber (2009). Media sentiment analysis based on big data has the potential to provide new opportunities for linking perceived brand image to customers (Zauner *et al*, 2012) and for further connecting the brand-related aspects (Echchakoui, 2018; Nam *et al*, 2017), but the analysis of how this should be done is left for other researchers to clarify. This study is also in line with Lipsman *et al* (2012), in understanding that media presence can be optimised by brand owners and companies. A new contribution is provided by introducing new types of analytics into the brand context to support these considerations.

5.1 Managerial implications and limitations

The managerial implications of this study include introducing a modern big data-based approach to brand image analytics. The practitioners can benefit from understanding the possibilities of media sentiment-based analytics, the current potential and limitations. It would be realistically possible to calculate and maintain the type of index introduced in this study. Those involved in marketing analytics might benefit from the understanding of the potential of analysing media sentiments. This study presents a possible new application. The approach of this study is oriented towards being practical. The implications that relate to the applied research method include demonstrating the utilisation of data mining and sentiment classification in brand-related research and analytics. This might be a growing research area in the future. Global opinion mining-based brand image was introduced as a concept.

The *limitations* of this study include analysing the potential of big data-based approaches for brand image analytics by using certain software to analyse media sentiment. The classifications are limited to the keywords used. Utilising different software might provide slightly different results due to different algorithms. Detailed content analysis of the data is not possible due to the sheer volume of the analysed data. The software also sets some limitations in terms of accuracy. The available computational capacity is also a limitation. The analysed time period was intentionally selected as rather short. Carrying out analysis on different time periods might provide slightly different results. The study does not take a position on brand-related cause and effect logic.

This study does not introduce a formula for creating a big data-based index on brand image but introduces possible logic and demonstrates the potential.

6 Conclusion

This study indicates that sentiment analysis based on big data has potential for analysing and calculating brand image indices. The main benefits of an automated brand index based on opinion mining involve new possibilities for analytics and continuous updating. The discussed index is a potential meaningful application of big data analytics. This study demonstrates the potential but does not introduce a formula for creating a big data-based index of brand image.

References

Ahmed, S., and A. Danti. 2015. A novel approach for sentimental analysis and opinion mining based on SentiWordNet using web data. Proceedings of the Trends in Automation, Communications and Computing Technology (I-TACT-15); 21-22 December 2015, Bangalore, India, pp.1-5.

Ball, L., and J. Elworthy. 2014. Fake or real? The computational detection of online deceptive text. *Journal of Marketing Analytics*, 2 (3): 187–201.

Barich, H., and P. Kotler. 1991. A framework for marketing image management. *MIT Sloan Management Review*, 32 (2): 94-104.

Brakus, J.J., B.H. Schmitt, and L. Zarantonello. 2009. Brand experience: What is it? how is it measured? does it affect loyalty? *Journal of Marketing,* 73 (3): 52-68.

Bruhn, M., V. Schoenmueller, and D.B. Schäfer. 2012. Are social media replacing traditional media in terms of brand equity creation? *Management Research Review*, 35 (9): 770-790.

Chakraborty, U., and S. Bhat. 2018. The effects of credible online reviews on brand equity dimensions and its consequence on consumer behavior. *Journal of Promotion Management*, 24 (1): 57-82.

Chan, T.K.H., X. Zheng, C.M.K. Cheung, M.K.O. Lee, and Z.W.Y. Lee. 2014. Antecedents and consequences of customer engagement in online brand communities. *Journal of Marketing Analytics*, 2 (2): 81–97.

Chaudhary, M., H. Kumar, S. Kaushal,and A.K. Sangaiah. 2018. The case analysis on sentiment based ranking of nodes in social media space. *Multimedia Tools and Applications*, 7 (4): 4217-4236.

Chaudhuri, A., and M.B. Holbrook. 2001. The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty. *Journal of Marketing*, 65 (2): 81–93.

Chauhan, K. and A. Pillai. 2013. Role of content strategy in social media brand communities: a case of higher education institutes in India. *Journal of Product & Brand Management*, 22 (1): 40-51.

Chen, H., R.H. Chang, and V.C. Storey. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36 (4):1165-1188.

Chen, L., 2018. Multilingual semantics applied to competitive and market intelligence. Proceedings of ICI 2018 international conference on Competitive & Market intelligence; 5-8 June 2018, Bad Neuheim, Germany.

Christodoulides, G., and L. De Chernatony. 2010. Consumer-based brand equity conceptualization and measurement: A literature review. *International Journal of Research in Marketing,* 52 (1), 43-66.

Deshmukh, J.S., and A.K. Tripathy. 2018. Entropy based classifier for cross-domain opinion mining. *Applied Computing and Informatics,* 14 (1): 55-64.

de Vries, L., S. Gensler, and P.S.H. Leeflang. 2012. Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26 (2): 83-91.

Dhaoui, C., C.M. Webster, and L.P. Tan. 2017. Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, 34 (6): 480-488.

Echchakoui, S. 2018. An analytical model that links customer-perceived value and competitive strategies. *Journal of Market Analytics*, 6 (4): 138-149.

Forbes. 2019. *Forbes most valuable brand list*. <Available online> https://www.forbes.com/powerful-brands/list/. Accessed 30.3.2019.

Fulgoni, G.M. 2018. Will digital commerce and analytics be the death of traditional brands? *Journal of Advertising Research, 58* (2): 146-150.

Goh, K.-Y., C.-S. Heng, and Z. Lin. 2013. Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*, 24 (1): 88–107.

Lee, T.Y., and R.T. Bradlow. 2011. Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48 (5): 881-894.

Jara, M., and G. Cliquet. 2012. Retail brand equity: Conceptualization and measurement. *Journal of Retailing and Consumer Services,* 19 (1): 140-149.

Jones, R. 2005. Finding sources of brand value: Developing a stakeholder model of brand equity. *Journal of Brand Management*, 13 (1): 10-32

Hillert, A., H. Jacobs, and S. Müller. 2018. Journalist disagreement. *Journal of Financial Markets*, 41: 57-76.

Hofacker, C.F., and D. Belanche. 2016. Eight social media challenges for marketing managers. *Spanish Journal of Marketing*, 20 (2): 73-80.

Krishen, A.S. and M. Petrescu. 2019. Data-driven decision making: implementing analytics to transform academic culture. *Journal of Marketing Analytics*, 7 (2): 51–53.

Kumar A J., S., Abirami, and T.E. Trueman. 2017. Sentiment Mining Approaches for Big Data Classification and Clustering. In *Modern Technologies for Big Data Classification and Clustering,* ed. H. Seetha, M. Murty, and B.K. Tripathy, 1-30. EBSCO Publishing.

Laroche, M., M.R. Habibi, M. Richard, and R. Sankaranarayanan. 2012. The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior, 28* (5): 1755-1767.

Lipsman, A., G. Mudd, M. Rich, and S. Bruich. 2012. The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising Research, 52* (1): 40-52.

Loughran, T. and B. McDonald. 2016. Textual analysis in accounting and finance: A survey, *Journal of Accounting Research*, 54: 1187-1230

M-Brain. 2015. *Corporate Communications*, received 11/2015.

McCorkindale, T., M.W. DiStaso, and C. Carroll. 2013. The power of social media and its influence on corporate reputation. *The Handbook of Communication and Corporate Reputation*, 497-512. Wiley.

MSG. 2016. *Management study guide*. <Available online> https://www.managementstudyguide.com/brand-image.htm. Accessed 18.2.2016.

Michaelidou, N., N.T. Siamagka, and G. Christodoulides. 2011. Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands. *Industrial Marketing Management*, 40 (7): 1153-1159.

Murphy, A. 2015. *2015 Global 2000: Methodology*. Forbes. <Available online> https://www.forbes.com/sites/andreamurphy/2015/05/06/2015-global-2000-methodology/#1faa616970f9. Accessed June 2019.

Nam, H., Y.V. Joshi, and P.K. Kannan. 2017. Harvesting Brand Information from Social Tags. *Journal of Marketing,* 81 (4): 88-108.

Nimmo, D., and R. Savage. 1976. *Candidates and their images: Concepts, attitudes and findings*. Santa Monica.

Nuortimo, K. and J. Harkonen. 2019. Improving strategic decision making with big data-based media analysis – The case of coal power. *International Journal of Global Environmental Issues*, In press.

Nuortimo, K., J. Harkonen, and E. Karvonen. 2017. Exploring the global media image of solar power. *Renewable & Sustainable Energy Reviews*, 81 (2): 2806-2811.

Olaleye, S.A., I.T. Sanusi, and J. Salo. 2018. Sentiment analysis of social commerce: a harbinger of online reputation management. *International Journal of Electronic Business*, 14 (2): 85-102.

Rohm, A., V.D. Kaltcheva, and G.R. Milne. 2013. A mixed-method approach to examining brand-consumer interactions driven by social media. *Journal of Research in Interactive Marketing*, 7 (4): 295-311.

Salinas, G., and T. Ambler. 2009. A taxonomy of brand valuation practice: Methodologies and purposes. *Journal of Brand Management*, 17 (1): 39-61.

Start-up. 2016. *The hottest startups in Finland*. <Available online> https://startup100.net/. Accessed 25.1.2016.

Swoboda, B., J., Weindel, and F. Hälsig. 2016. Predictors and effects of retail brand equity–A cross-sectoral analysis. *Journal of Retailing and Consumer Services*, 31: 265-276.

Wang, W., L. Chen, K. Thirunarayan, and A.P. Sheth. 2012. Harnessing twitter 'Big Data' for automatic emotion identification. in Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on Social Computing (SocialCom), IEEE, Washington, DC. 587-592.

Weinberg, B.D., L. Davis, and P.D. Berger. 2013. Perspectives on Big Data, *Journal of Marketing Analytics*, 1 (4):187-201.

Xu, X., G.L. Frankwick, and E. Ramirez. 2016. Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69 (5): 1562-1566.

Zauner, A., M., Koller, and M. Fink. 2012. Sponsoring, brand value and social media. *Revista De Administração De Empresas, 52* (6): 681-691.
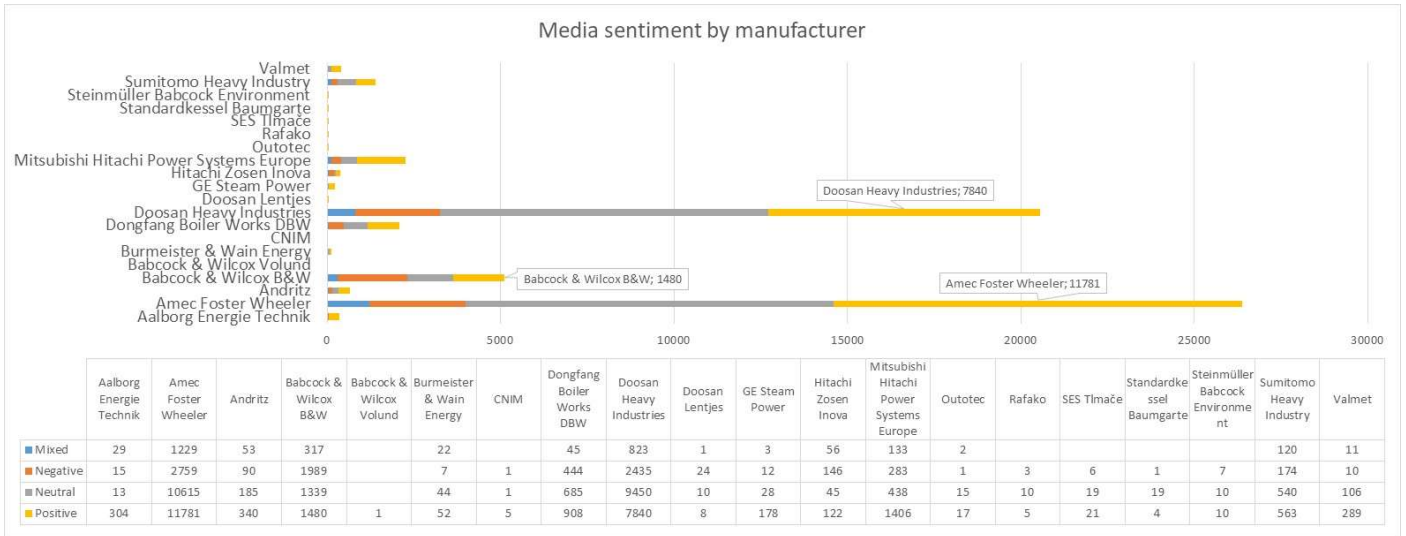
Figure 1. Media sentiment of power equipment manufacturers

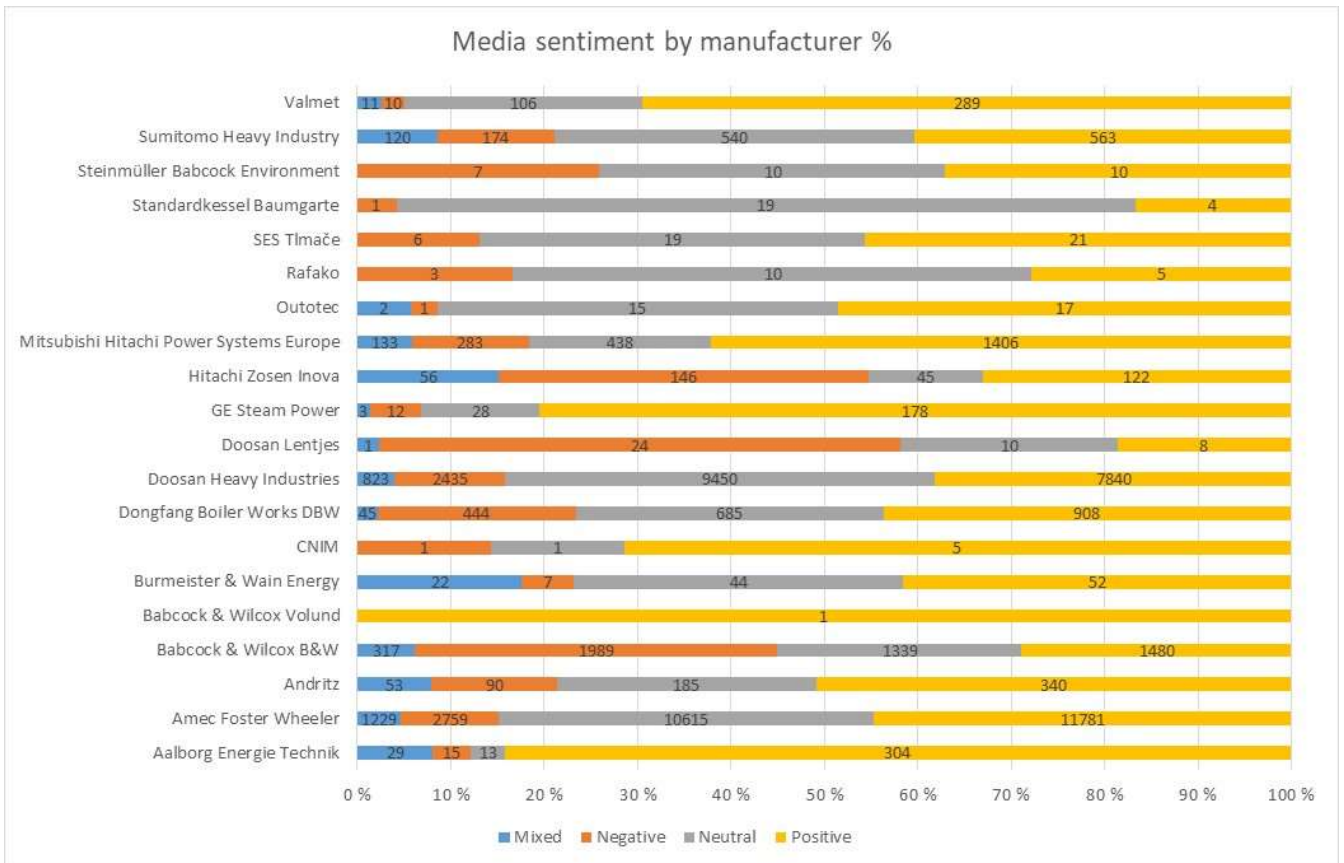|	Aalborg Energie Technik	Amec Foster Wheeler	Andritz	Babcock & Wilcox B&W	Babcock & Wilcox Volund	Burmeister & Wain Energy	CNIM	Dongfang Boiler Works DBW	Doosan Heavy Industries	Doosan Lentjes	GE Steam Power	Hitachi Zosen Inova	Mitsubishi Hitachi Power Systems Europe	Outotec	Rafako	SES Tlmače	Standardkessel Baumgarte	Steinmüller Babcock Environment	Sumitomo Heavy Industry	Valmet
Mixed	29	1229	53	317		22		45	823	1	3	56	133	2					120	11
Negative	15	2759	90	1989		7	1	444	2435	24	12	146	283	1	3	6	1	7	174	10
Neutral	13	10615	185	1339		44	1	685	9450	10	28	45	438	15	10	19	19	10	540	106
Positive	304	11781	340	1480	1	52	5	908	7840	8	178	122	1406	17	5	21	4	10	563	289

Figure 2. Percentual relative sentiment classification of media hits related to power equipment manufacturers

Table 1. Forbes brand index compared to opinion mining

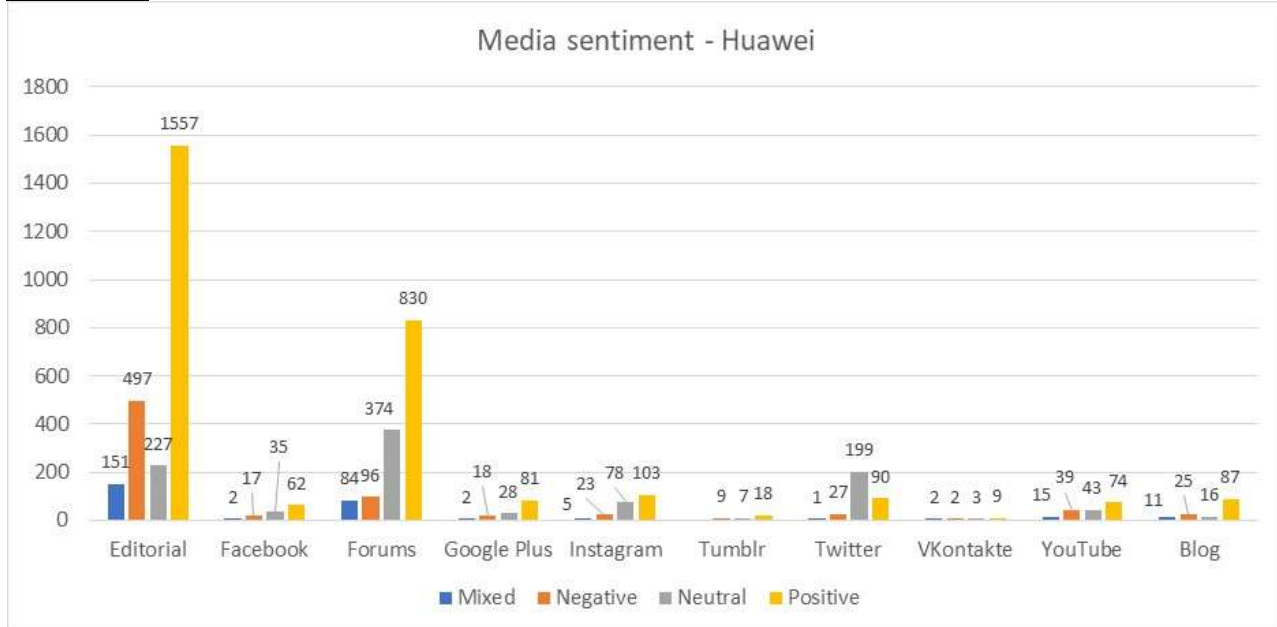| Company	Forbes brand index		Data from opinion mining (one month)		
	Position on Forbes list	Brand value	Total media hits	Total hits: % negative	Social media hits: % negative
Huawei	88	$7.3 B	4959	15 %	29.87 %
Adidas	75	$7.9 B	4920	14.98 %	20.89 %

Figure 3. One-month global media visibility with sentiment classification - Huawei
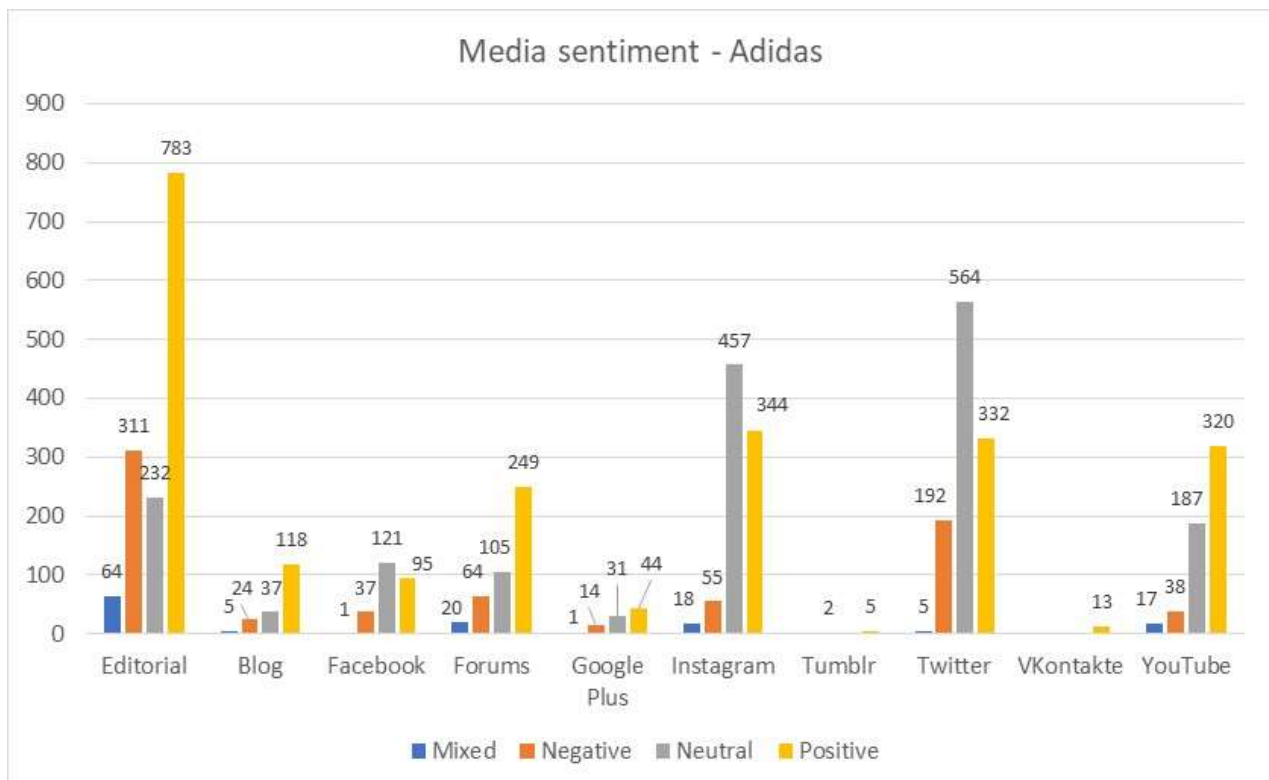

Figure 4. One-month global media visibility with sentiment classification – Adidas

Table 3. Proposed brand image index structure

Brand	Media sentiment (One year)			Company rank (1–100)
	Total media hits	% negative	% negative on social media	
Brand 1				
Brand 2				
Brand 3				
Brand 4				
Brand n				