The predictive power of Social Media Analytics: To what extent can SM Analytics techniques be classified as reliable and valid predictive tools?

Author: Nicolas Philipp Kalmer
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT
This critical literature review investigates three promising Forecasting / Social Media Analytics techniques: (1) trend analysis, (2) topic modeling and (3) sentiment analysis. To evaluate these methods, several assessment criteria are derived from Hammann & Erichson's (1994) criteria of market segmentation and prognosis. In total, there are five criteria which have to be taken into consideration, the criterion of observation, the criterion of measurement, the stability of the dataset, the purchasing relevance, and the criterion of homogeneity. To further question these forecasting methods in terms of methodological composition, several concepts of Babbie (2009) (validity & reliability) are used. Eventually, it is the goal of this critical evaluation to determine the predictive and analytic power of these forecasting methods resulting in a methodological taxonomy. In addition to this, the value of these techniques for Marketing strategy is questioned.
In essence, it can be concluded that a large amount of managerial practitioners and scholars prognosticated Social Media and Big Data to facilitate Marketing activities such as market segmentation or the forecasting of customer behavior. However, it has to be emphasized that forecasting within social media environments is a sensitive undertaking bearing a large amount of potential research biases, may it be viewed from the quality of data or the sophistication of the forecasting method. This outcome calls for further research in the realm of Social Media Forecasting due to the fact that best forecasting practices have not been established yet.

Supervisors: 1st DR. A.H. (RIK) VAN REEKUM
2nd DR. E. (EFTHYMIOS) CONSTANTINIDES

Keywords
Social Media, Big Data, Social Media Analytics, Forecasting Reliability / Validity, Forecasting Taxonomy, Marketing Strategy

6th IBA Bachelor Thesis Conference, November 5th , 2015, Enschede, The Netherlands.

1. INTRODUCTION

The 21st century with its accelerating technological spirit and interface has been door-opening for a large amount of new product developments. Inter alia the internet has brought along Social Media, which can be acknowledged as a sub-category of Big Data platforms, and can be considered as the most remarkable innovation due to the fact that respective role models such as LinkedIn, Facebook, Twitter, Instagram and youTube, to name a few, have penetrated everyday life to large extent and have become an essential part of private and professional life. In a nutshell, right now the professional landscape is at the beginning of the Big Data era and more and more ventures are in the process of "capitalizing" these large datasets, or the knowledge gained from these (Fania & Miller, 2012).

By having a closer look at the professional/business dimension, it is self-evident that especially the Marketing activities have been revolutionized by the advent of Social Media and the increased connectivity and online presence of customers. With respect to the aforementioned movements, the Internet of Things (increased amount of mobile, connected and portable devices) and Web 2,0 (emergence of user-generated content), the possibilities Social Media offer seem to be endless (Constantinides & Fountain, 2008; Del Chiappa et al., 2014). As a matter of fact, ventures aim to jump on this hype train and participate actively, due to the fact that proper Data Analysis ("Social Media Analytics", "Data Mining") provides the venture with "deeper and richer insights into business patterns and trends, help drive operational efficiencies and competitive advantage in manufacturing, security, marketing, and IT" (Fania & Miller, 2012). The list of studies and respective advantages about Social Media Analytics, Data Mining and SM Prediction/Forecasting power is endless. Nonetheless, numerous scholars claim that there is a research gap concerning a classification of the most promising SM analytic techniques (Asur & Huberman, 2010; Schoen et al., 2013). Schoen et al. (2013) argue that there is no common ground regarding the predictive power of Big Data due to the fact that a comparative classification of forecasting methods in a SM setting has not been established yet. The study further claims that most models and techniques used to investigate past regularities are solely applicable in their respective setting, whereas it remains questionable whether these reveal external validation. Recently, it has become of utmost importance in academic research to identify a forecasting method, which may be acknowledged as the holy grail of predictive modeling, to be used in multiple scenarios, in order to facilitate the forecasting procedures of proactive ventures. This investigation can be considered as worthwhile with respect to the wide and fragmented range of forecasting methodology (Asur & Huberman, 2010; Schoen et al., 2013; Fania & Miller, 2012).

Nonetheless, there is an agreement in academia about the research potential of Big Social Data. It may be considered as a virtual playground for researchers of multiple domains, delivering a vast amount of cheap and real-time customer data offering a large amount of exploration opportunities (Schoen et al., 2013; Asur & Huberman, 2010; Fania & Miller, 2012). In addition to this, Fania & Miller (2012) encourage further investigation of these datasets due to the multifarious range of advantages, such as the quality of information (unstructured, but first-hand customer information), the acquisition costs (extremely low), the innovative input (possible co-creation of new product developments), the exploration of market needs, the explosive nature (spread, reach and information dissemination). Moreover, SM analytics are in the position to be used to further strengthen customer relationship management (He et al., 2013; Fan & Gordon, 2014) or to further customize NPDs and to conduct Marketing activities in an efficient and targeted manner (Kosinski et al., 2013).

However, one also has to take a look at the other side of the coin. Numerous authors, such as Couper (2013), question the value of Big Social Data and reveal a large amount of potential research biases, such as demographic biases, measurement pitfalls, data manipulation or the file-drawer effect. Preliminary, it can be acknowledged that the conceptions towards the usability of Big Social Data vary to large extent, bearing both exploration opportunities and bias potential. Next to the quality of the data gathered, this research will reveal particular interest in the design of the forecasting methods, due to the fact that, regardless of the data quality, a sophisticated forecasting method may be in the position to capture and analyze the meaningful data components. In sum, it has to be emphasized that a large amount of studies disregard the sensitivity of Big Social Data Forecasting (Couper, 2013; Wijnhoven & Bloemen, 2014; Pang & Lee, 2008). Thus, this study will examine the SM analytic techniques fastidiously in terms of data capturing and data analysis in order to cover and question the potential pitfall areas described by Couper (2013).

2. RESEARCH AGENDA & METHODOLOGY

This study will be conducted by means of a critical literature review. It aims to investigate the forecasting potential of Social Media Analytic techniques in terms of criteria of market segmentation and prognosis in combination with the issue of reliability and validity. Whereas prior studies have already revealed the predictive power of particular big data-based methods in their respective setting, a critical, methodological taxonomy of the various forecasting techniques is still lacking. The demand for this taxonomy is of utmost importance due to the fact that the current forecasting work is widely fragmented (Kalampokis et al., 2013). In addition to this, researchers may be misguided by the file-drawer effect, indicating that their model claims to reveal universal applicability, whereas it remains questionable in how far these methods score in terms of external validation and reliability (Couper, 2013). As a consequence, this study critically questions three of the most promising forecasting methods in accordance with prognosticating criteria and methodological concepts. With respect to the aforementioned thoughts, the following main research question can be derived:

- To what extent can Social Media Analytics techniques be classified as reliable and valid predictive tools?

As the research question already implies, the scope of analysis will be focusing on both the data's perspective in terms of quality and the methods themselves in terms of reliable and valid predictive architecture. Moreover, several sub-questions will be introduced to further enlighten the misty nature of this topic. The first one represents the main part of research

- What types of SM Analytic techniques have been established yet?

and examines the most promising SM Analytic techniques by conducting a critical evaluation of these in accordance with several evaluative criteria derived from the well-known German economist Peter Hammann, who wrote a book about the issue of market research and prognosis. In addition to this, further sub-questions will be answered.

- What is the "added value" of SM Analytics?
- What value does SM Analytics provide with respect to Marketing Strategy?

Eventually, the taxonomy of methods creates a multi-dimensional analysis of the presented forecasting methods. To further broaden the scope of analysis, the overall potential of these methods is going to be assessed by having a closer look at the "added value" of SM forecasting and the impact to Marketing strategy.

Concerning the literature review, academic databases such as Scopus or Google Scholar have been used as information sources in line with scientific journals such as the journal of Technological Forecasting and Social Change. Key search terms have been "Social Media Analytics", "Social Media Prediction", or "Forecasting and Social Media". In the course of time method specific cues, such as "Trend Analysis", "Twitter Analysis", "Topic Modeling" or "Sentiment Mining", have been used.

3. KEY CONCEPTS

3.1 Social Media

With respect to the aforementioned thoughts, some crucial key explanations will follow in order to familiarize the readership with the basic contextual definitions. It is of utmost importance to outline these concepts ahead of the initial analysis with respect to the unique and sensitive process of Social Media data gathering and analysis. The reason for concept selection is closely adjusted to the main research question, which calls for in-depth explanations of Social Media, Social Media Analytics and criteria which may be used to assess reliability and validity.

The study of Zeng et al. (2010) classifies social media to be a "controversial, distributed mode of content generation, dissemination, and communication among communities". In addition to this, the authors emphasize the key feature of Social Media, indicating that the "information consumption and dissemination process is becoming intrinsically intertwined with the process of generating and sharing information". Eventually, it is emphasized that weblogs, microblogs, online forums, wikis, podcasts, live streams, social bookmarks, web communities, social networking, and avatar-based virtual reality may be considered as Social Media. Kietzmann et al. (2011) promote the honeycomb-framework of SM, which classifies SM in seven functional building blocks - presence (the awareness of other users available), sharing (information exchange frequency, information dissemination), relationships, online identity, conversations/communication, reputation (standing of other participants) and groups (within the community). All these dimensions require specific attention by the network host. In sum, Kietzmann et al. (2011) claim SM to "employ mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content".

Taking both definitions together, one may derive certain "character traits" of SM, which can be summarized as controversial discussions/content/information, large spread, huge information dissemination, communication between individuals, consumption and creation of content, information sharing, online, "artificial" identities, and reputation.

All the aspects need to be considered in detail when it comes to the analysis of data within online environments. The reasons for this consideration are multifarious. For instance, in case online attention is measured as presented in Asur & Huberman (2010), one cannot simply relate the box office success of a movie to the amount of Tweets the movie received in advance. One may also consider opinions/sentiments, the amount of Tweets per user, the type of Tweet (visual/textual/nonsense) or the proportion of Re-Tweets concerning a particular movie. Moreover, the SM definition is of utmost importance by having a closer look at the words used in the Tweet or the fact that "satisfied" consumers/possible movie watchers tend to be silent in online environments (Asur & Huberman, 2010). With respect to the aforementioned thought "words used in the Tweet", one has to be careful when it comes to the Tweet database extraction. In particular, Asur & Huberman had to erase certain movies from their investigation list (for instance the blockbuster "2012") for the simple reason that this movie title could be easily used in a completely different context leading to biased predictions. This short example should sharpen the reader's awareness of the difficulties when it comes to the data collection and the associated limitations in online environments.

3.2 Social Media Analytics

There are numerous definitions of Social Media Analytics. According to Zeng et al. (2010) SMA "is concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyse, summarize, and visualize social media data", eventually applied to "extract useful patterns and intelligence" within the dataset. SMA being a research area in its own, having gained increased interest in the past years by academics, but also by practitioners and ventures which specialise in SMA and offer tools and services to other businesses and individuals, outline today's significance of this area. Fan & Gordon (2014) outline SMA's value with respect to a lifecycle analysis. In the "design-development" stage, for instance, SMA can be helpful to analyse trends in tastes, to promote product innovation and improvements by "capturing and understanding conversations involving either of two groups". Further in-depth skills of SMA will be presented in the respective sub-question investigating the added value of SMA.

3.3 Social Media Analytics evaluation criteria framework

In the following section several forecasting methods evaluation criteria will be introduced. Inter alia they belong to the wider content of market research and market analysis and are derived from market segmentation criteria (Hammann & Erichson, 1994) due to the fact that, in essence, from a certain angle, forecasting methods can be understood as market segmentation attempts in order to discover customer needs and market trends based on patterns detected in data. The idea is to further name and explain five essential evaluation criteria, which are combined in the position to test whether a forecasting method can be considered to have a sound predictive architecture and value for marketing strategy. Previously, some of these criteria have been used by Hammann & Erichson (1994), who used them to segment markets. In this case, it is assumed that Social Media Analytics can be interpreted as means of market segmentation and prognosis.

The criteria have been selected due to the fact that they are in the position to assess a forecasting method's value according to multiple domains. The criteria of observation, measurement and stability cover the method's evaluation in terms of validity and reliability, whereas the purchasing relevance and the homogeneity enlighten the readership about the method's value for marketing strategy. Thus, taking all the different criteria into account, they are combined in the position to question a forecasting method from both angles, the methodological composition and the 'real-life' value. Hammann & Erichson (1994) developed these criteria long before the rise of the internet, so it remains questionable whether these are also applicable for the online analysis. However, it seems likely that this transmission should not be considered to be problematic due to the fact that SM environments are often assessed as societal mirrors representing the collective wisdom of the society (Schoen et al., 2013)

3.3.1 The criterion of Observation

The first criterion is about the observability of the dataset. For instance, concerning a Twitter analysis, it goes without saying that there is a large amount of attributes that one may observe, such as possible key words extracted by crawling mechanisms, hashtags, links to other Twitter users or different types of emoticons used. All of these aspects are relatively easy to observe. However, there may be some aspects which cannot be observed at first sight, such as irony / sarcasm in Tweets (appearance of Trolling) or hidden messages influencing the Tweet's sentiment. The question is whether a method is indeed in the position to deal with these hidden considerations.

Another difficulty is for instance Instagram, as a non-text based form of SM which cannot be easily measured as a text based example where words can be counted.

3.3.2 The criterion of Measurement

The second criterion is related to the measurement. To keep the aforementioned example of a Twitter analysis, one may measure a large amount of particular aspects about Tweets, for instance Average-Tweet-Rates per hour per movie, such as Asur & Huberman (2010) did, to forecast box-office revenues. One may even measure the amount of Tweets per user or the amount of Re-Tweets concerning a certain issue. However, there are certain things which are rather difficult to measure, such as the impact of visual content via URLs, videos or movie trailers. It may be a severe difficulty to determine the return on visual investment due to the fact that often those visual posts are less frequently used, they may outperform non-visual Tweets though concerning their impact as these invite the user to interact.

3.3.3 The criterion of Purchasing Relevance

The third criterion is related to the purchasing relevance.

In essence, firms use their data mining techniques to detect patterns and trends in data and aim to capitalize the knowledge gathered from these. The interesting question is how companies get the user from like to buy. What is the purchasing relevance of this particular method? Are users only creating attention and modern gossip or are they really intending to buy something? Users may tweet frequently about a particular movie. However, it is of utmost importance to determine whether the interested crowd is going to pursue the purchase in the end. The user's interest in a particular product may be seen as valid to make a purchase, but is it really reliable in the end?

Related to the movie example, one may question whether it is actually ensured that the users are going to watch the movie in the end, or whether they intend to get the movie using illegal ways.

Long before the rise of the internet professional marketers had to deal with the problems of predictability and prognosis and developed tools, which over decades were widely accepted by advertising agencies and the industry. The most common criterion for the success of advertising (ads in magazines or spots in television) has been the so called recall. Peer groups were shown several TV commercials and were asked which they did remember. The assumption of the entire marketing world was the higher the recall value the more sales would happen. It was advertising legend David Ogilvy (1983), founder and long-time chairman of the world-wide advertising network Ogilvy, now being a part of London based WPP group, who shattered this myth in his famous book about advertising by mentioning that for gaining a high recall value a gorilla should be equipped with suspenders (Ogilvy, 1983).

Science always relies on meaningful descriptions of cause and effect. It remains a secret what generations of advertisers have believed that recall of ads and TV commercials could prove a logical result in terms of real sales.

Making the transition to the forecasting methods evaluation, it will be critically assessed whether these methods can be considered as purchasing-irrelevant or reveal at least a slight monetary impact since this is what Marketing is all about.

3.3.4 The criterion of Stability

The fourth criterion requires a certain kind of stability of the dataset. Especially in online environments opinions may shift within seconds due to the real-time nature. Again, SNS tend to be relatively instable environments, inter alia due to the fact that the majority of users uses these platforms to file complaints about certain issues rather than expressing positive feelings (Asur & Huberman, 2010). Therefore, several SM based platforms tend to be easily biased since the satisfied users prefer to be silent as they have no grievances to complain about.

3.3.5 The criterion of Homogeneity

The fifth criterion is the criterion of homogeneity. It claims that a forecasting method should result in a market segment which is more homogeneous than the total market. For instance, in case one uses Twitter data as forecasting input, one may be aware of the fact that the data observed is demographically biased. However, this can either be beneficial or counter-productive, depending on the investigated topic.

4. SM ANALYTIC TECHNIQUES

4.1 Trend Analysis

Trend analysis is one of the most common forecasting methods in online environments due to the fact that this statistical approach, entailing procedures such as time-series analysis or linear regression, to name a few, aims to detect past regularities within historic datasets to make accurate predictions about the future, and Social Media deliver these datasets at large scale. In a nutshell, the amount of studies using a regression analysis to forecast particular events is endless (Asur & Huberman, 2010; Achrekar et al., 2011). Hence one of the most promising studies will be presented in detail. Asur & Huberman (2010) developed a regression model to make accurate predictions about the box-office revenues of particular movies, based on the "attention" it previously received on the well-known micro-blog Twitter. In essence, the authors created a predictive regression model incorporating three different variables. Firstly, the authors considered the amount of attention, a particular movie received prior to its release measured in Tweet traffic. To further operationalize this Tweet traffic concerning a particular movie, the authors introduced the Average-Tweet-Rate (ATR) per movie, which is calculated by dividing:

(Tweets per movie / Time in hours) to receive the average tweet rate per hour.

The study assumes that movies with a high ATR will receive a relatively high revenue in the end.

And indeed, quite surprisingly the simple relationship between the ATR and the box-office revenue for the 24 movies analysed revealed a relatively strong correlation ($R=.90$; Adjusted R-square=.80). Nonetheless, the authors aimed to further optimize their forecasting approach.

Secondly, the authors conducted sentiment analyses concerning the extracted Tweets in order to categorize the user's opinion /

sentiment (positive;negative;neutral). The sentiment assignment had been conducted by the Amazon Mechanical Turk (AMT), who attached the sentiments to the respective Tweets in order to create a training dataset for the used language classifier. Eventually, the classifier revealed 98% accuracy.

The sentiment analysis' impact had been only marginal though.

Thirdly, the authors considered the "distribution" of the movies which represents nothing else than the amount of theatres the movie had been shown in. This consideration revealed an even stronger statistical correlation with an Adjusted R-square of 0.973 for the ATR-theatre count in relation to the revenue.

Eventually, Asur & Huberman came up with the following regression model:

*y = ßa*A + ßp*P + ßd*D + e*

y=revenue to be predicted

ß=respective values according to the regression coefficients

A=attention rate (ATR)

P=polarity of the sentiments (pos. sentiments / neg. sentiments)

D=distribution (related to theatre count)

e=error

4.1.1 Criticism on Asur & Huberman's Twitter Forecasting

The model itself seems to be quite promising and advanced by having a closer look at the presented results. The simple correlation between the ATR and the respective box-office revenue outperformed other big players within the forecasting domain such as the Hollywood Stock Exchange (HSX) and the model continuously improved its accuracy the more variables have been taken into account. According to Asur & Huberman (2010) the model may be applied within other research domains without any severe difficulties.

Nevertheless, it remains questionable whether Twitter Data is reliable to be used as forecasting input. The authors claim Twitter Data to be a reliable dataset representing the "collective wisdom" of the society (Asur & Huberman, 2010). However, when dealing with data collection in online environments there will always be a large amount of concerns and limitations, such as unpredictable user sampling compositions (for instance age biases), fake identities misleading research, or analysis restrictions concerning non-viral topics (Yu et al.,2012). Yu et al. (2012) summarize the most important aspects of successful Twitter prediction: topic of interest has to be human-thought-related; user composition needs to be sound; issue should not be humiliating to tweet about in order to invite a large amount of users to share their point of view.

Asur & Huberman (2010) criticise other studies within their research domain for focusing on non-representative data such as the movies' genre or the amount of famous actors within the cast as these only entail the audiences' general perception towards these aspects, whereas Asur & Huberman (2010) rely on real-time data only and consider the importance of virality.

Nonetheless, the authors made a controversial move within the capturing phase of the CUP model by Fan & Gordon due to the fact that they used the Twitter Search API tool (non-commercial Tweet extractor) for Tweet extraction, which is only in the position to extract an amount of Tweets which is lower than 1% of the total Tweets (Wang et al., 2012).

Wang et al. (2012) criticise this Search API and doubt its reliability, therefore the authors used a commercial data extractor in their study to capture an appropriate amount of Tweets. Surprisingly, the study of Asur & Huberman (2010) revealed relatively accurate results. However, one has to bear in mind that the prediction of movie revenues may be an exceptional case due to its viral character. One may conclude that the analysis of niche topics within Twitter is too difficult to execute due to the fact that the low amount of Tweets does not justify to generalize the outcomes.

In addition, Asur & Huberman (2010) had to erase some movies from their investigation list since there have been blockbusters such as "2012", of which the title may have been used in a completely different context leading to biased results. Moreover, when dealing with Twitter data one has to consider the proportion of Re-Tweets concerning the datasets and the amount of Tweets per single user as these aspects may cause biased results as well.

The analysis will be structured as follows: At first, the method will be analysed and assessed according to the previously introduced criteria of market research and prognosis by Hammann & Erichson (1994). In addition, the criteria will be translated into methodological concepts and considerations by Earl Babbie (2009), in case the criteria allow to be scientifically mirrored revealing reasonable methodological counterparts.

Eventually, concluding remarks will follow to summarize the predictive value of this forecasting method by having a closer look on its dealing with crucial methodological concepts such as validity and reliability of data.

(1) Criterion of Observability: When conducting a Twitter Analysis, there is always a certain difficulty involved concerning the observability of the units of analysis, i.e. the Tweets or particular attributes of interest. In general, Babbie (2009) distinguishes between direct observables, such as word count, emoticons used or hashtags, and indirect observables, such as sentiment polarity or irony / sarcasm. It seems to be self-evident that studies about direct observables are favored by researchers due to the fact that these are easier to deal with in terms of observation and measurement. However, when dealing with high polarity data such as Tweets, especially the indirect observables shape the researcher's interest, due to the fact that a large amount of business functions, especially the Marketing & Sales department, is highly dependent on these opinions of customers towards, for instance, product features. By having a closer look at Asur & Huberman's (2010) regression equation, one can detect three different variables, the Average-Tweet-Rate (ATR), which depicts the average Tweets per movie per hour, the sentiment, revealing the user's opinion towards a particular movie (positive;negative;neutral), and finally the theatre count, which can be easily assigned to the movies and requires no further investigation due to its stable and external nature.

Asur & Huberman (2010) deal with the ATR calculation and the sentiment assignment by using the Twitter Search API as the Tweet extraction tool, which is free of charge. Its value for academic investigation remains questionable and will be addressed in the upcoming section "stability of the dataset". In essence, a large amount of authors claim that this tool is not representative at all due to the fact that it only extracts less than 1% of the total amount of Tweets. For instance, the study of Wang et al. (2012) relies on a commercial extractor which allows to capture a wider range of Tweets. Nevertheless, for now this issue should no longer be questioned. Next, the authors calculated the ATR per movie per hour, which can be easily observed and calculated. Thus, concerning the ATR calculation no bias can be detected. Similar results can be acknowledged for the sentiment variable due to the fact that the authors consulted workers of Amazon Mechanical Turk (AMT) for the sentiment assignment with at least three workers per single Tweet to determine the correct polarity. So, concerning the observation itself no severe mistake has been done, enabling the regression equation to produce reliable outcomes for different input. Throughout the upcoming sections it will be questioned though to what extent this procedure proves to be valid.

To further increase the readership's awareness of grievances concerning Twitter research and the data collection / observation in online environments in general, the study of Wang et al. (2012) proposes several aspects to consider:

(1) The aspect of virality: The authors claim that there is a topic restriction to conduct analyses via Twitter since some topics may reveal a lack of public interest which consequently results in a shortage of Tweets.

(2) Additionally, one may take into account that some issues could be humiliating to Tweet about which may also prohibit the users to express their thoughts. However, this problem does not apply to Asur & Huberman's (2010) study due to the fact that movies can be considered as a viral topic.

(2) Criterion of Measurement: To further investigate the predictive architecture of the forecasting method, it needs to be questioned in terms of validity and reliability. Babbie (2009) distinguishes between four types of validity (face-validity, criterion-related validity, construct validity, content validity). The author defines validity as follows: "A term describing a measure that accurately reflects the concept it is intended to measure". To provide an example, Babbie mentions that "your IQ would seem a more valid measure of your intelligence than the number of hours you spend in the library would". According to the author this concept is of utmost importance to assess the overall value of a research project. He further emphasizes that validity must not be confused with the term reliability, due to the fact that reliability questions whether a particular method is able to generate the same consecutive results multiple times when it is applied within the same context (Babbie, 2009). These terms will now be mirrored concerning Asur & Huberman's regression analysis. At first, the analysis will be questioned in terms of its face-validity. According to Babbie (2009) face validity refers to "that quality of an indicator that makes it seem a reasonable measure of some variable". To provide the readership with an example, the author mentions the frequency of attendance at religious services as an appropriate indicator of the person's religiosity. It can be concluded that the study of Asur & Huberman has face-validity with respect to the assumption that the Average-Tweet-Rate (ATR) can be seen as an appropriate indicator of the variable "attention on Twitter". This conception requires no further explanation due to its basis on common sense. The other variables are self-explanatory and do not require further investigation.

Moreover, the method has to be analysed in terms of Criterion-related validity. Babbie (2009) emphasizes that this concept deals with the ability of the measure to forecast a variable which may be seen as an external standard which is automatically increasing the studies' value by reaching this level. And indeed, the predictive architecture of Asur & Huberman's (2010) study seems to be outstanding, since their model is in the position to outperform other well-known industry metrics, such as the Hollywood Stock Exchange (HSX), which may be seen as the holy grail in the realm of prediction and forecasting.

Surprisingly, the simple correlation between the ATR and the box-office revenue for the 24 movies considered revealed a relatively strong correlation ($R=.90$; Adjusted R-square$=.80$). The other variables further increased the model's accuracy, especially the sentiment component had only a marginal impact though.

Furthermore, the technique is questioned in terms of its construct validity. Babbie (2009) defines construct validity as "the degree to which a measure relates to other variables as expected within a system of theoretical relationships". Babbie (2009) mentions the example of measuring marital satisfaction, which's construct validity is, for instance, further strengthened by the finding that satisfied husbands and wives are less likely going to be unfaithful than dissatisfied ones. However, it can be acknowledged that most of the predictive models investigated fail to explain the techniques by the construct of theories (Asur & Huberman, 2010; Achrekar et al., 2011). Ultimately, a method reveals an appropriate construct validity in case theoretical constructs are in the position to represent real-life situations they are intended to model (Babbie, 2009). However, most of the predictive methods do not elaborate on the logical and causal relationships amongst the variables.

Eventually, the method will be questioned in terms of content validity. According to Babbie (2009) content validity refers to "the degree to which a measure covers the range of meanings included within a concept". The author emphasizes that, for instance, a test of mathematical ability must not be restricted to addition, it needs to cover subtraction, multiplication, division and other components. Transferred to Asur & Huberman (2010) the main variable investigated is online attention. Indeed, the study seems to reveal a solid content validity due to the fact that it also investigates the polarity of the attention or the type of material (for instance visual content) which may have caused the respective online chatter.

In essence, it can be concluded that most of the predictive studies fail to develop other supporting measures concerning the construct validity, which can be related in a logical way to the established measures and provide the construct with solid evidence.

(3) Criterion of Stability of the dataset: The next criterion is dedicated to the quality and stability of the dataset. In addition, the data is going to be questioned in terms of its reliability. Babbie (2009) defines reliability as "that quality of measurement method that suggests that the same data would have been collected each time in repeated observations of the same phenomenon". In other words, the concept of reliability investigates whether a certain research technique generates the same results over and over again. In general, within a SM setting the affirmation of reliability is rather difficult to secure due to the fact that a large amount of SM applications reveal a real-time nature, indicating that the data input changes every single time the method is applied yielding different outcomes every single time. Couper (2013) further introduces several difficulties concerning the real-time data collection in online environments, which will be addressed later on in the respective sub-question (see added value of SMA). However, Babbie (2009) further points out that the aspect of reliability does not incorporate accuracy. Especially online environments may reveal several obstacles to accuracy, such as demographic biases, sampling biases, data manipulation or fake identities (Couper, 2013). Generally speaking, it can be concluded that Big Data offers a large amount of research opportunities due to data overload, one trade-off of this Big Data era is definitely the accuracy of data though. With respect to the aforementioned thoughts, several biases have already been discussed. Babbie (2009) further emphasizes some areas of potential biases, such as the area of observation. For instance, it may be difficult for researchers investigating employee behaviour, to determine whether workers are kidding with each other or having an argument. This issue may further be of difficulty in case there is only one observer ("single observation"). At least for the sentiment component, Asur &

Huberman (2010) dealt with this issue quite well due to their decision to let the Tweets be assigned by at least 3 employees to prohibit this observation bias. Moreover, the issue of reliability questions the usage of data sources, so whether the research is based on a single data source or multiple ones. The application of several validity measures such as the test-retest method can only be conducted in case there are at least two data sources. However, most predictive studies fail to include a second data collection domain due to the fact that one has to deal with domain specific phenomena only. In sum, SM data usage bears a large amount of potential bias foundations. These will further be mentioned in the course of the paper.

(4) Criterion of purchasing relevance: Viewed from a Marketing angle, the purchasing relevance should be seen as the most important criterion of forecasting methods due to the fact that ultimately Marketing is all about the sales impact. However, it remains questionable whether Asur & Huberman's regression analysis can be assigned with this categorical criterion. Nevertheless, it cannot be denied that this type of analysis bears a large amount of favorable indicators concerning the eventual sales volume, such as a huge ATR per movie, positive sentiment polarity or a large theatre count. Despite all the evidence to the contrary, it cannot be guaranteed that the movie will be well-watched in the end. This problem is ubiquitous in forecasting literature since the majority of studies presents a large amount of indicators of the ultimate outcomes (sales volume), it cannot be ensured though. However, the regression analysis of Asur & Huberman may be seen as an advanced sales volume predictor due to the fact that the model outperformed other well-known industry predictive metrics (HSX). Thus, due to the model's solid criterion-related validity, it may be assumed that the sales volume predictions can be indeed accurate.

(5) Criterion of Homogeneity: The study of Asur & Huberman handles the issue of homogeneity quite well. In the end, the predicted results can be related to a market segment which is more homogeneous than the initial market and consists of active Twitter users interested in movie releases. Therefore, the output can be associated as very helpful for further targeted Marketing activities due to the fact that it represents a relatively small niche market segment.

4.2 Topic Modeling

The next Social Media Analytics technique which will be introduced is the topic modeling method. Fan & Gordon (2014) emphasize that "topic modeling uses a variety of advanced statistics and machine-learning techniques; for instance, a number of models identify "latent" topics through the co-occurence frequencies of words within a single communication or between topics and communities of users". Next to the described latent topic identification, Wallach (2006) points out that it is the technique's core purpose to "identify representations of the data that reduce description length and reveal inter- or intra-document statistical structure. So, it can be concluded that topic modeling deals with the identification of topics within textual data, which are in the position to represent the data's content based on statistical classification.

Blei (2013) further emphasizes the demand for advanced statistical measures and algorithms to deal with large bodies of text due to the fact that recently there has been a shift towards neurotic digitalization concerning the storage of any type of knowledge and data. In addition, Blei (2013) defines topic modeling as a "suite of algorithms that aim to discover and annotate large archives of documents with thematic information". The author heavily criticizes the current browsing and information gathering process via online research since regular search engines detect only separate documents relating to the keywords searched for, whereas topic modeling enables researchers to gather thematically connected documents even with respect to the thematic change in the course of time. Moreover, the author introduces the most applied topic modeling method, which is called Latent Dirichlet Allocation (LDA: “The distribution that is used to draw the per-document topic distributions is called a Dirichlet distribution”), which is based on a rather complex algorithm, the process itself can be transferred to a simplified example though. When this LDA topic model is applied to textual data it works comparable to a set of pens, which are highlighting related words within the investigated document with different colors, making it easier for researchers to detect the hidden themes within this textual document, eventually. The main advantage of this algorithm compared to other statistical classifiers is that it handles extremely large bodies of text automatically which require no prior labeling or tokenization, which can be understood as a pre-processing data technique which is often needed to remove noisy data to make the input classifier-fitting (Wang et al., 2012). So, the basic idea behind the LDA method is that a textual document consists of a large amount of sub-categories and topics, which are statistically distributed over a previously fixed amount of vocabulary. With respect to the aforementioned thought, it is crucial to emphasize that every document theoretically incorporates the same themes, however with "different proportions of exhibition" (Blei, 2013). In practice it can be concluded that this model observes words in documents to identify hidden variables such as the underlying topic structure or the generative process with which the document has been designed. Eventually, it is the goal of the algorithm to calculate the joint probability distribution concerning observed and hidden data (also called posterior distribution) based on the statistical distribution of the hidden variables given the observed ones.

(1) Observability: As previously introduced, when it comes to the observability of data within online environments, one has to distinguish between direct and indirect observables (Babbie, 2009). Concerning the LDA algorithm, comparable issues as examined concerning the regression analysis, need to be taken into consideration. Apparently, the algorithm is in the position to observe thematic information (topics and sub-categories) within large bodies of textual data. Even though this sophisticated algorithm seems to be able to handle extremely large collections of documents and even whole archives, the method itself reveals some restrictions concerning skills which go beyond this rather simple topic assignment. It has to be emphasized that the algorithm has not been trained to handle sentiment and opinion polarity within documents due to its descriptive nature. Indeed, the algorithm reveals no further analytic potential which could possibly be investigating issues such as sentiment polarity, author's subjectivity, irony / sarcasm or idiosyncrasies in language. Among other things, these issues will be further taken into consideration when it comes to the stability criterion and the detection of potential biases.

(2) Measurement: Again, the LDA method will be mirrored towards Babbie's concept of validity. However, this time it seems reasonable to narrow the scope of validity to a certain degree to facilitate the readership's awareness of the essential validity issue regarding Blei's topic model. The model may be questioned in terms of content-validity, since with respect to the aforementioned thoughts, it remains questionable whether it is sufficient to analyse these archives solely in terms of content. Indeed, it seems to be justified to further question the model's dealing with semantic structures or meta-language, which will most likely not be part of the previously assigned word treasure. Nevertheless, especially for Marketing activities the detection of the author's signature within these large bodies

of textual data seems to be required. In addition to this, the model may be questioned in terms of face-validity. According to Blei (2013) the different topics incorporate a certain word treasure with single words being displayed with a certain probability within this particular topic. It is further emphasized that this distribution is created in advance and does not change in the course of time. Thus, the model's face-validity is highly dependent on these word treasure probability distributions, which may be inaccurate due to the fact that these are not further displayed in the study.

(3) Stability: From a certain perspective, one may argue that topic modeling deals with stable datasets due to the fact that this method has been particularly designed to increase the browsing experience through large bodies of text and even whole archives. So, it can be concluded that this method involves no real-time dimension concerning the data gathering as opposed to Twitter research. However, there are still certain grievances which may hinder the reliability of the outcomes. For instance, the continuous development of speech challenges the worthiness of this algorithm since the method only deals with a fixed word treasure foundation which requires further adjustments and may be outdated in the near future. In addition to the development of speech which is a rather slow process, there may be some other aspects challenging the algorithm's reliability, such as language idiosyncrasies and slang vocabulary. It cannot be assumed that the algorithm is trained to handle these types of data, thus there is a clear restriction in terms of applicability which prohibits the algorithm to be used in environments which are known for a certain vernacular and informalities. Moreover, the extent to which the algorithm covers meta language and professional vocabulary remains questionable. Nevertheless, the algorithm can be seen as a sophisticated data processing method for textual data, whereas it should focus on historic datasets to determine developments concerning certain themes rather than exploring real-time based communities or documents displaying sentiment polarity and idiosyncrasies.

(4) Purchasing: The concept of purchasing relevance does not entirely apply to this method due to the fact that users are known for expressing their intention towards a certain product and the possibly resulting purchase in micro-blogs such as Facebook or Twitter, in which the algorithm does not reveal validation since no training concerning the online vernacular has been incorporated in its design. However, the algorithm may be suitable for product review blogs as these incorporate longer and rather professional documents which may be worth it to be properly investigated. Nonetheless, the direct relation of purchasing to this type of modeling still remains unclear. In any case, a forecasting method displaying the final purchasing volume of customers could be considered as the holy grail for Marketers due to its clear depiction of the relation between investment and return on investment. One may argue that Asur & Huberman's study presents the answer to this question. However, the outcomes are clearly topic-related and require a certain virality concerning the investigated issue.

(5) Homogeneity: The criterion of homogeneity cannot be applied to topic modeling due to the fact that topic modeling is restricted to discovering topics and sub-categories in data rather than dealing with the segmentation of markets.

4.3 Sentiment Analysis

Both terms, opinion mining and sentiment analysis, will be used interchangeably during this essay and are concerned with the application of "computational linguistics, natural language processing and other methods of text analytics to automatically extract user sentiments or opinions from text sources at any level of granularity (words or phrases up to entire documents)" (Fan & Gordon, 2014). In a broader setting, the term itself has gained increasing attention by scholars and managerial practitioners in accordance with the emergence of Online Social Networks (OSNs) which have been designed under the premise to display the opinions and sentiments of the users towards a specific event, happening, development or product (Brun, 2011). With respect to the aforementioned thoughts, recently researchers have revealed particular interest in micro-blogs such as Twitter or review blogs and forums such as Amazon due to the fact that these platforms incorporate relatively short messages which a huge degree of polarity to current discussions or events, presenting a fruitful foundation of academic investigation (Krauss et al., 2008; Bollen et al., 2011). However, one has to bear in mind that a large amount of surveys are written by Big Data Analytics fanatics, who propagate their analytic methods to be universal and groundbreaking. It remains questionable though whether this is indeed the case.

An interesting approach to track the sentiments of posts / Tweets, is provided by Wang et al. (2012) who created a real-time analysis framework to analyse the public's mood, expressed in Tweets, towards the presidential candidates of the US elections in 2012. The authors justify their research investigation with the claim that the emergence of real-time data collection mechanisms "vastly outpaces the capacity of traditional content analysis approaches, calling for novel computational approaches" (Wang et al., 2012).

In detail, the authors extracted and analysed approximately 36 million Tweets via the commercial Twitter application tool "GNIP power track" which collects the Tweets and searches for Tweet specific features such as hashtags used with respect to specific political candidates, idiosyncratic expressions, sentiments expressed by emoticons, URLs, Retweets, links to other users, repetitions and sarcasm / irony. In total, this approach may be considered as quite advanced compared to other studies with respect to the fact that the researchers examined about 200 possible features of Tweets for an amount of 9 political candidates, the above mentioned included to name the most important ones. In addition to the "sentiment-carrying" emoticon extraction, the authors consulted Amazon Mechanical Turk (AMT) to let the Tweets' sentiment be assessed on a manual basis by the consulted group of annotators. The annotators had to assess the Tweet's sentiment (positive;negative;neutral;unsure) and attach additional features such as the question whether a Tweet could be considered as sarcastic / humorous. Eventually, the authors used a statistical sentiment classifier performing 59% accuracy with the AMT assessment.

(1) Observability (idiosyncratic expressions ,hidden messages): The authors emphasize the extreme difficulty to collect and observe data in online environments. Most of the data input is of terrible quality and gathered in informal networks, in which one has to deal with a certain type of online network "vernacular" and idiosyncratic expressions. In addition, it is pointed out that traditional content analysis approaches are no longer applicable to these types of idiosyncrasies, "calling for novel computational approaches" to deal with these. However, the authors miss to name examples of traditional content analysis approaches in line with their respective inabilities to solve this language processing problem. In any case, one may suppose that the authors handle the issue of observability quite well due to the incorporation of about 200 tokenizers tracking possible language idiosyncrasies.

(2) Measurement: However, the statistical classifier used (based on naive Bayes classification) seems to be quite advanced due to the fact that it covers a wide range of tokenizers to correctly classify idiosyncrasies, emoticons, URLs, Re-Tweets, user links, hashtags, repetitions, sarcasm,

phone numbers, and so forth. Nonetheless, it needs to be criticized that this language classifier performs only 59% accuracy, which can be acknowledged as an average result.

(3) Purchasing Relevance: The authors developed a real-time sentiment framework which may be transferred to multiple domains. In addition to the sentiment about politicians, it could possibly reveal the sentiment towards a company's ad campaign, TV commercial or product feature. However, it clearly remains difficult to relate the sentiment polarity to the eventual sales impact. Nonetheless, this advanced type of classifier can only be the answer to the network specific idiosyncrasies.

(4) Stability of Dataset: The Twitter data source cannot be considered as stable due to its real-time nature.

(5) Homogeneity: Does not apply.

5. ADDED VALUE OF SMA

The rise of Web 2.0 and Social Media applications and platforms in its many forms in the last decade, driven by technological advances with the internet at its core, created various opportunities from a business standpoint, for both application and research. At the centre of it stands the vast amount of data, which is accumulated and spread by Social Media in general and the users' discussions, tweets, posts, blogs and videos as forms of collaboration and interaction in particular. The according buzzword is termed 'Big Data' and describes "a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques" (Webopedia). More fitting in the context of this paper, it will be referred to 'Big Social Data', being the amount of data generated by Social Media websites and applications. The central question then evolves around how to manage and analyse these massive datasets and capture value for one's company, which brings us to the field of Social Media Analytics (SMA). SMA is "concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application" (Zeng, Chen, Lusch & Li, 2010). SMA being a research area in its own, having gained increased interest in the past years by academics, but also by practitioners and ventures which specialise in SMA and offer tools and services to other businesses and individuals, outline today's significance of this area.

It can be capitalised on SMA in several areas. First of all from a research perspective, Big Social Data can be understood as data springing from a "living-lab" in a "real-word environment", as Stieglitz, Dang-Xuan, Bruns and Neuberger (2014) point out. Manipulating and analysing this data gives researchers a valid alternative to conventional data gathering methods such as surveys. Next to that, the global reach of Social Media allows researchers to derive at insights about different groups of people regarding culture and ethnicity for instance, as well as allowing for comparative studies of e.g. certain events in different regions. The technological advance of today largely diminishes the constraints of distance and the availability and accessibility to information (based on Big Social Data) can facilitate research in a lot of areas, provided that it is properly processed by means of SMA tools and methods. From a practical perspective then, SMA finds its use to enable companies to specifically target advertising, manage PR and social customer relationships and most importantly aid in utilising social media as an additional channel for marketing (Stieglitz et al., 2014). SMA as an element of integrated marketing strategy will be assessed in more detail in a following part. In this direction, Zeng et al. (2010) put forward that besides the for-profit sector benefitting from SMA, it is also public entities which are beginning to embrace the potential SMA can offer. Political parties, government agencies such as homeland security and public health institutions can profit from Big Social Data. To be more precise, political parties and individuals for example monitor public opinion on certain issues, identify trending topics and build community support for their agendas. Government agencies naturally tap into social media and its data pool for security reasons and terrorism prevention. In a recent publication by Fan and Gordon (2014), the authors outline the business value of SMA by means of a lifecycle analysis. In the early 'design-development' stage, SMA can be employed analyse trends in tastes and promote product innovation and improvements by "capturing and understanding conversations involving either of two groups". It further helps to forecast demands, mitigate production risks and generally improve the supply chain responsiveness ('product production stage). The most common use of SMA however takes place in the 'product utilization' stage. In this stage it is sought to analyse customer buying behaviour and tastes to come up with an overall better customer segmentation. The three key social media objectives as identified by Hoffman and Fodor (2010), brand awareness, brand engagement and word of mouth, are paramount factors for most marketing departments and of high priority when analysing social media. Finally, Fan & Gordon (2014) also highlight the significance of SMA with its potential of identifying and responding to crises.

While the above paragraph outlines the value SMA can add in today's business setting, one has to be careful not to overlook the challenges coming along with it. First and foremost, the social media environment is fairly young and very volatile meaning that now established baselines may be obsolete or not adequate anymore as new platforms, objectives and also analytical methods can surface quickly.

Other challenges of Big (Social) Data in general and SMA in particular are potential bias in form of "selection bias and self-presentation (measurement) bias", the aforementioned volatility and lack of stability (rather short-term trends than long-run measurement), privacy and access issues, "opportunity of mischief" as in non-reliable data due to the anonymity of the internet and the sheer mass of data, which is by nature very hard to manage and interpret. (Couper, 2013) To be added to that, is the importance to choose the appropriate method or approach in order to make sense of the data, of where there are several in the SMA realm. Also the objective of one's initiative has to be clearly determined.

6. FORECASTING TAXONOMY

Three promising forecasting methods have been presented as exemplary studies of their respective domain. In essence, it needs to be emphasized that these methods generate reasonable results in their respective setting, whereas a lack of external validation can be acknowledged due to multifarious reasons. The regression analysis of Asur & Huberman (2010) does not reveal a large amount of grievances concerning the concepts of validity and reliability though. Especially in terms of content and criterion-related validity the model can be acknowledged as an outstanding predictive method. Outperforming external criteria such as the HSX and examining not only quantitative factors of analysis (ATR) but also qualitative factors (sentiment mining and theatre count) the method can be considered as an advanced and sophisticated predictive means. The accuracy of the results further confirms this evaluation. Nonetheless, the method has a weakness concerning the construct validity due to the fact that this method is not based on theoretical constructs of cause and effect. However, it can be emphasized that the majority of trend analysis models do not focus on this dimension and fail to establish logical relationships among the variables. In addition to this, the method provides a huge impact for Marketers due to the fact that this technique is in the position to accurately predict concrete numerical values, which can be seen as a unique feature and of utmost importance for Marketing since it relates to the return on investment. However, one has to be aware of the data sources one is using when conducting predictive analytics. Twitter, Facebook and other real-time micro-blogs reveal a large amount of potential research biases related to the concept of reliability, whereas it can be pointed out that reliability does not ensure the accuracy of a predictive method (Babbie, 2009). The sentiment analysis of Wang et al. (2012) is exactly confronted with this issue of accuracy. Even though the authors established a sentiment classifier based on naïve Bayes Classification, carrying about 200 (!) language specific tokenizers, it could only ensure 59% accuracy. This result can be associated as average. It seems that even though the classifier has been well-tokenized there is still a mismatch between the idiosyncrasies used in Tweets and the previously classified tokenizers. Indeed, it seems that especially micro-blogs such as Twitter and Facebook incorporate a certain vernacular, which is rather difficult to deal with. This calls for novel approaches in sentiment mining with respect to the continuously evolving language, meta language and semantic structures. An elegant but rather time consuming method to compensate the classifier's inaccuracies seems to be the consulting of sentiment annotators who assign the Tweet's sentiment in small groups to ensure accurate outcomes. Blei's (2013) topic modeling method can also be acknowledged as an advanced analytical technique due to its ability to structure large archives of textual data and assign them with topics and respective sub-categories. It goes without saying that this method has to be examined in a non-social media based setting since it is not applicable to short messages carrying a certain vernacular and polarity. In any case, this method can be used in multiple domains due to the fact that it is capable of structuring, for instance, the whole history of a certain topic and how it changed in the course of time. This skillful algorithm enables Marketers to browse online environments at large scale identifying crucial thematic information within seconds, based on the previously assigned word treasure distributions. However, the analysis fails to explain how these word distributions are derived. It can only be assumed that these compositions are based on historic statistics, which may be challenged in the course of time with respect to evolving themes and according specified vocabulary and content which has not been classified yet. In any case, it should have become clear that SMA is only in the beginning of its development. Right now, researchers are fanatic about the research exploration opportunities of Social Media and micro-blogs and develop predictive methods to specific scenarios, which is the right approach since these multi-dimensional OSNs incorporate a large amount of complex considerations which cannot be handled by a single method with equal distribution. The visualized taxonomy can be found in the appendix section (Table1).

7. DISCUSSION & VALUE FOR MARKETING STRATEGY

According to Hammann & Erichson (1994) the marketing strategies' effectiveness and success is related and highly dependent on the activities of market segmentation and prognosis. In detail, the authors emphasize that market segmentation and prognosis are conducted by marketers to adjust the ventures undertakings to the differentiated needs of the consumers. In sum, the authors define two main purposes of market segmentation and prognosis, which are the analysis, indicating that the customers of a total market ought to be classified in homogeneous groups, and strategy development, aiming at the selection of appropriate market segments which meet the ventures' capabilities and skills. So, related to the previously introduced criteria of market segmentation and prognosis, the criteria of homogeneity, purchasing relevance and stability seem to be of specific interest concerning the assessment of the value for marketing strategy. In other words, a method can be considered of having a high value for marketing strategy in case the technique creates a homogeneous market segment which is more stable than the total market and reveals a certain degree of purchasing relevance. Taking the exemplary studies of Asur & Huberman and Wang et al. into account, one may argue that these studies successfully segment markets due to the fact that these analyse Tweets which are related, inter alia through hashtags, to a very specific theme only. However, it needs to be emphasized that this type of market segmentation has to be seen very critically since one does not get to know other attributes about the user other than that he/she is an active Twitter enthusiast and is somewhat involved in the particular topic. The detection of these attributes cannot be considered as an advanced market segmentation technique, due to the fact that crucial knowledge about the end-user is missing, which can be seen as essential for efficient targeting. In any case, the sophistication of the market segmentation is obviously dependent on the type of marketing strategy, which may be serving niche markets or rather heterogeneous markets, which do not require advanced segmentation. In addition to this reasoning, it is self-evident that topic modeling is not related to the issue of market segmentation due to the fact that it deals with the discovery of topics and sub-categories within textual data rather than the identification of user segments. Nevertheless, besides from the single criterion of homogeneity, methods can prove their value for marketing strategy in other categories. For instance, the topic modeling technique may be used within forums or extensive product review blogs to detect user's attitudes towards product features and similar issues. In any case, one has to be aware that the dealing with Social Media data prohibits the researcher in the majority of cases to get to know the user/consumer in-depth due to the fact that online users tend not to share a large amount of details within their profile, and even if they did, it remains questionable whether this additional input may serve as a fruitful foundation of further market segmentation and targeting. This is a good time to mention the 'traditional' survey model, which raises specific questions towards the respondents which go beyond the expression of polarity and short messages. For an efficient and targeted marketing campaign one should rather rely on sophisticated market segmentation techniques, which self-evidently are rather expensive, however they are expected to manage the issue of segmentation and prognosis in a reliable manner (Hammann & Erichson, 1994). In any case, some specific topics may not require advanced segmentation techniques, in case the targeting in conducted in a very broad manner and the market tends to be rather heterogeneous.

8. CONCLUSION

The analysis of SMA techniques along with several criteria of market segmentation and prognosis and the concepts of validity and reliability has shown that forecasting in SM settings is a highly sensitive undertaking. In essence, due to multifarious potential research biases, SM data cannot be considered as a reliable and accurate data source. Future researchers have to be aware of a critical mindset towards SM data and this study delivers several methodological concepts and evaluative criteria, which may be used as a starting point of future critical evaluation of forecasting methods and related data sources. Nonetheless, with respect to the presented forecasting taxonomy, each single exemplary study reveals a large amount of benefits and favorable indicators within the applied context, whereas most of the well-known forecasting methods fail to reveal a certain degree of external validation. However, initially it has been assumed that there may a predictive model which may be applied within different scenarios, but this would only weaken the method's value due to the fact that it attempts to generalize forecasting and SM predictive means, which is a foolish assumption with respect to the conception that every single forecasting scenario requires a unique treatment. Future research may bear these concluding remarks in mind. When properly executed, SMA is in the position to deliver great value for marketing strategy and the business in general.

9. ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. van Reekum for his helpful guidance and provision of valuable thematic information during my Bachelor assignment. In addition, I would like to thank Dr. Constantinides for his effort as a second supervisor.

10. REFERENCES

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S., & Liu, B. (2011). Predicting flu trends using Twitter data. Computer Communications Workshops (INFOCOMWKSHPS), 2011 IEEE Conference, 702-707.

Asur, S., & Huberman, B. (2010). Predicting the Future with Social Media. International Conference On Web Intelligence And Intelligent Agent Technology, 01, 492-499.

Babbie, E. (2009). The practice of social research (13th ed.). Belmont, Calif.: Wadsworth Pub. Co.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal Of Computational Science, 2(1), 1-8. http://dx.doi.org/10.1016/j.jocs.2010.12.007

Brun, C. (2011). Detecting Opinions Using Deep Syntactic Analysis.

Constantinides, E., & Fountain, S. (2008). Web 2.0: Conceptual foundations and marketing issues. J Direct Data Digit Mark Pract, 9(3), 231-244. http://dx.doi.org/10.1057/palgrave.dddmp.4350098

Couper,. (2013). Is the sky falling? New technology, changing media, and the future of surveys. Paper Presented At The Survey Research Methods., 145-156.

Del Chiappa, G., Lorenzo-Romero, C., & Constantinides, E. (2014). Disintermediation and User-generated Content: A Latent Segmentation Analysis. Procedia - Social And Behavioral Sciences, 148, 524-532. http://dx.doi.org/10.1016/j.sbspro.2014.07.075

Fan, W., & Gordon, M. (2014). The power of social media analytics. Communications Of The ACM, 57(6), 74-81. http://dx.doi.org/10.1145/2602574

Fania, M., & Miller, J. (2012). Mining Big Data in the Enterprise for Better Business Intelligence. IT@Intel White Paper.

Goel, S., Hofman, J., Lahaie, S., Pennock, D., & Watts, D. (2010). Predicting consumer behavior with Web search. Proceedings Of The National Academy Of Sciences, 107(41), 17486-17490. http://dx.doi.org/10.1073/pnas.1005962107

Hammann, P., & Erichson, B. (1994). Marktforschung (3rd ed.). Stuttgart: Gustav Fischer.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. International Journal Of Information Management, 33(3), 464-472. http://dx.doi.org/10.1016/j.ijinfomgt.2013.01.001

Hoffman, D., & Fodor, M. (2010). Can You Measure the ROI of Your Social Media Marketing?. Magazine: Fall 2010 Research Feature.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. Internet Research, 23(5), 544-559. http://dx.doi.org/10.1108/intr-06-2012-0114

Kietzmann, J., Hermkens, K., McCarthy, I., & Silvestre, B. (2011). Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons, 54(3), 241-251. http://dx.doi.org/10.1016/j.bushor.2011.01.005

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings Of The National Academy Of Sciences, 110(15), 5802-5805. http://dx.doi.org/10.1073/pnas.1218772110

Krauss, J., Nann, S., Simon, D., & Fischbach, K. (2008). Predicting movie success and academy awards through sentiment and social network analysis.

Ogilvy, D. (1983). Ogilvy on advertising. New York: Crown.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations And Trends In Information Retrieval, (2(1-2), 1-135.

Schoen, H., Gayo-Avello, D., Takis Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. Internet Research, 23(5), 528-543. http://dx.doi.org/10.1108/intr-06-2013-0115

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social Media Analytics - An interidsciplinary Approach and its Implications for Information Systems. Business & Information Systems Engineering, (Vol. 6: Iss. 2), 89-96.

Wallach, H. (2006). Topic Modeling: Beyond Bag of Words. Proceedings Of The 23Rd International Conference On Machine Learning.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. ACL System Demonstrations, 115-120.

Wijnhoven, F., & Bloemen, O. (2014). External validity of sentiment mining reports: Can current methods identify demographic biases, event biases, and manipulation of reviews?. Decision Support Systems, 59, 262-273. http://dx.doi.org/10.1016/j.dss.2013.12.005

Yu, S., & Kak, S. (2012). A survey of prediction using social media. Arxiv Preprint Arxiv:1203.1647.

Zeng, D., Chen, H., Lusch, R., & Li, S. (2010). Social Media Analytics and Intelligence. IEEE Intell. Syst., 25(6), 13-16. http://dx.doi.org/10.1109/mis.2010.151

11. APPENDIX

Forecasting Taxonomy

	Methodological Concept	Trend Analysis (Asur & Huberman, 2010)	Topic Modeling (Blei, 2013)	Sentiment Analysis (Wang et al., 2012)
Observation	Direct & Indirect Observables	Outstanding dealing with the issue of online observation due to the inclusion of 'hidden variables' such as sophisticated sentiment analysis & inclusion of external 'real-life' variables	Restricted to larger bodies of textual data, not applicable to Micro-blogs or OSNs; skilled algorithm in the position to observe topics & sub-categories based on previously designed word treasures; not able to capture vernacular & evolving language	Actually well-trained and tokenized statistical classifier with about 200 idiosyncrasies; fails to provide outstanding accuracy; fails to deal with short messages carrying certain vernacular and polarity
Measurement	Validity	Outstanding validity coverage in terms of face, criterion & content validity; method fails to provide construct validity due to the fact that it is not based on theoretical constructs	Model lacks content validity due to the fact that semantic structures & meta-language cannot be analysed by this method; analysis is restricted to the word treasures & word probability distributions	Sentiment Analysis is highly dependent on tokenization & training of the classifier; classifier fails to generate accurate results even though there has been a large amount of pre-processing activities
Purchasing	/	Method does not ensure sales, it delivers a large amount of sales-related indicators though to assume this; Twitter Sales Prediction may be restricted to viral topics only	Has no direct purchasing relation, the method may be applied to review blogs and forums though in case the entries are algorithm-fitting in terms of length and language	Good method to capture real-time data, it remains questionable though whether there is a clear relation between the Tweet's sentiment and the sales impact, it can rather be seen as a favorable indicator
Stability	Reliability, Accuracy & Research Biases	Stability of Twitter data is not given due to its real-time nature; several biases are taken into account without further checking; no re-test using a different data source; neither reliable nor accurate	There may be an inaccuracy due to the fact that the probability distributions are not explained in-depth by the authors; continuous development of speech may cause analysis restriction (also: meta-language & vernacular)	Stable environment is not given due to Twitter's real-time nature
Homogeneity	/	Handles issue of homogeneity quite well, in the end the results are related to a market	/	/

		segment which is more homogeneous than the total market		
Conclusion	/	Advanced and differentiated approach of predictive modeling; huge value for Marketing Strategy; cannot be considered as entirely reliable & accurate though	Outstanding method to increase the browsing experience regarding topics & sub-categories; however highly dependent on the previously designed word treasures and the respective probability distributions; has problems with meta-language & SM data	Well-trained sentiment classifier, generates average results, reveals that sophisticated sentiment analysis requires further data pre-processing and tokenization

Table1. Forecasting Taxonomy of SMA