CORPORATE CREDIT RISK ANALYSIS UTILIZING TEXTUAL USER GENERATED CONTENT – A TWITTER BASED FEASABILITY STUDY

Aaron Mengelkamp, Chair of Application Systems and E-Business, University of Goettingen, Goettingen, Germany, amengel@uni-goettingen.de

Sebastian Hobert, Chair of Application Systems and E-Business, University of Goettingen, Goettingen, Germany, sebastian.hobert@uni-goettingen.de

Matthias Schumann, Chair of Application Systems and E-Business, University of Goettingen, Goettingen, Germany, mschuma1@uni-goettingen.de

Abstract

Irrecoverable receivables resulting from insolvent debtors endanger the own liquidity. Therefore, corporate credit risk analysis should be continuously improved in order to diminish bad debt. We analyse in how far user generated content (UGC) contains evidence concerning the financial stability of companies and hence, can enhance the information base for corporate credit risk analysis. For this purpose, we compare data from the microblogging platform Twitter related to ten insolvent and ten solvent German companies. We utilize techniques from content analysis for the quantification of textual data. Results from independent t-tests indicate, that the amount of UGC is significantly higher and the sentiment is significantly worse for insolvent companies prior to the date of bankruptcy than for solvent companies in the same time span. Furthermore, we apply the k-Nearest-Neighbour algorithm in order to classify companies as prospectively insolvent or solvent based on sentiment scores derived from UGC. Results show, that a classification accuracy above randomly expected values can be achieved. The classification accuracy increases, when UGC published closer to the date of insolvency is used. Future research should focus on how to utilize our findings and improve processes of corporate credit risk analysis while integrating UGC.

Keywords: Credit Risk, Text Mining, Twitter, User Generated Content.

1   INTRODUCTION

Bad debt can lead to illiquidity of businesses. Therefore, lenders avoid giving credit to debtors with a high probability of default. Hence, it is important to accurately assess the probability of default for clients when these purchase on account. Then, irrecoverable receivables caused by insolvent debtors can be prevented (Stiglitz & Weiss 1981). Furthermore, the debtor needs to be constantly monitored because credit agreements and associated default risks underlie constant changes especially in ongoing credit relationships (e. g. supplier-credit relationships) (Beck 2014).

Creditors use a broad basis of information and several screening devices in order to assess the credit worthiness of debtors and reduce the existing information asymmetry in this principal-agent relationship (Stiglitz 1988). This asymmetry of information consists of hidden action and hidden information. Hidden action refers to the circumstance that the creditor (principal) cannot monitor the actions the debtor (agent) has undertaken or will undertake in the future perfectly (e.g. invested effort in the going concern principle). Only the output (e. g. receipt of payment) can be observed. Hidden information is information (e. g. forecasts of the agents turnover and profit) which is available to the agent but the principal does not know about (Arrow 1985). The data used to reduce this information asymmetry is essentially different for private and commercial debtors. We focus on corporate credit risk analysis in this paper because the credit volume and hence the risk exposure is usually greater than for private debtors (Beck 2014). For commercial debtors the information base consists of internal information existing from previous credit relationships (e. g. payment behaviour) and external information (e.g. financial figures from annual reports or reports from credit agencies) which are updated at certain points of time only (Graham & Coyle 2000; Schumann 2002). For example, changes in payment behaviour can only be detected when new business transactions are finalized and financial figures can only be updated when new financial reports are available. Next to that, industrial, structural and management analyses are conducted (Chee et al. 1999; Graham & Coyle 2000). Long term characteristics such as strategic, operational and technological indicators for an impending insolvency are also monitored (Graham & Coyle 2000; Hall & Young 1991). Nevertheless, short term indicators of impending insolvencies especially for small firms such as environmental circumstances (e. g. theft or fire) are not monitored with the current data set (Hall & Young 1991). The data is analysed with statistical methods such as discriminant/regression analysis as well as machine learning algorithms like neural networks or the k-Nearest-Neighbour algorithm. Typical characteristics of good and bad debtors are identified utilizing these methods (Altman 1968; Galindo & Tamayo 2000). Then, the potential debtor is classified into one of the groups based on his characteristics. If these are similar to characteristics of bad debtors, the creditor will most likely renounce the credit relationship. The debtors are not categorized into two groups only in state of the art credit assessment. They are assigned to more elaborate rating categories which can implicate different interest rates or payment terms (Keßler 2010). The procedures for credit assessment should be continuously improved in order to further diminish bad debt (Kramer & Nitsch 2010).

Bartels (2013) proposes research endeavours in which the applicability of data from social media, which are "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (Kaplan & Haenlein 2010, p. 61) for improvements in credit assessment processes should be evaluated. Hereby, UGC "can be seen as the sum of all ways in which people make use of social media" (Kaplan & Haenlein 2010, p. 61). This means, that any content in social media which is created by end-users and not by the platform operators or the system itself is UGC (Kaplan & Haenlein 2010). It seems suitable to integrate textual UGC as an additional source of information in credit assessment because it is proven that early indicators for changes in economic and commercial conditions can be extracted from it (Bollen et al. 2011). Furthermore, UGC is updated more frequently and in real time in contrast to information sources used in state of the art credit risk analysis and hence, holds potentials to provide important information earlier (Bartels 2013; Kaplan & Haenlein 2010). Start-ups like Kabbage (2015) offer corporate credits based on a rating established on a company's online appearance and reviews e. g. from Ebay, Amazon or Yahoo. However, the

rating models of Kabbage are not revealed in order to protect the competitive advantage of Kabbage and their appropriateness has not been comprehended yet.

Therefore the research question we answer in this paper is formulated as:

In how far can textual UGC extend the information base for corporate credit risk analysis?

In the next section, we review literature concerning text mining in the financial domain. Then, we derive research hypotheses and describe our dataset consisting of 7071 Tweets related to 20 German companies. After that, our research setting as well as our data analysis methods which encompass manual content analysis and quantitative statistics are presented. Subsequently, we present our findings and discuss these. At the end, we reveal limitations of our research and conclude, which practical and theoretical implications arise from our results.

2 RELATED LITERATURE

In order to give the reader an overview on used methods and latest research results in text mining in the financial domain, we summarize existing literature in the following.

Sentiment scores derived from public corporate disclosures, news and textual UGC can be mapped with financial data in order to detect statistical coherences e. g. with firm fundamentals or for price prediction of financial instruments. Hereby, it can be distinguished between manual, dictionary and machine learning based approaches used for quantification of textual data (Kearney & Liu 2014; Nassirtoussi et al. 2014).

Hirokawa et al. (2011) solely rely on manual content analysis. They extract causes of bankruptcy from news and claim, that reasons for bankruptcy can often be found close to the word "but". In the other papers, manual classification is used to train a machine learning algorithm only, but the results of the automated classification results are not compared to manually classified data sets. However, it is suitable to start with manual content analysis especially when text analysis is performed in new contexts such as research domains, languages or UGC from diverse social media platforms, since it tends to be most accurate (Stieglitz et al. 2014).

In dictionary based approaches, algorithms map words, phrases or sentences with pre-defined dictionary categories. Scores are assigned to the dictionary categories in order to determine the sentiment of the textual documents. Hereby, the word list and the term weighting scheme are important. Most word lists do not include words which have a positive or negative meaning in the financial domain, but are created for the general English language (Kearney & Liu 2014). Context-specific word lists achieve better results than general dictionaries (Henry & Leone 2009). Furthermore, in most studies the words in the dictionary are treated as equally important and the sentiment score is calculated based on frequency counts of words appearing in the dictionary categories (Kearney & Liu 2014). However, recent results show, that an appropriate weighting of words in the dictionary is more important than a comprehensive word list. (Jegadeesh & Di Wu 2013). Furthermore, results indicate, that especially negative sentiment from textual data is connected to future firm fundamentals and market prices (Bollen et al. 2011; Kearney & Liu 2014; Nassirtoussi et al. 2014). Exceptional high and low pessimism leads to increased trading volumes. In addition to that, high pessimistic values cause increased market volatility (Kearney & Liu 2014). The only dictionary based work which can be directly related to corporate credit risk analysis is from Tsai et al. (2010). They analyse effects of news coverage and sentiment factors on credit ratings. They conclude, that firms with high news coverage are rated worse in the next quarter. This effect is stronger for firms with speculative grades. Furthermore, news polarity concerning firms with investment grades is connected to their future ratings.

When machine learning algorithms (e. g. Neural Networks or Naïve Bayes) are applied, a sample of data is manually classified (Nassirtoussi et al. 2014). This sample is then used to train an algorithm. Training means, that the algorithm creates a model based on the manually classified data set which contains the statistical rules for text classification. Before inserting unstructured textual data into a machine learning algorithm, it needs to be pre-processed. Hereby, features of the data need to be

selected. The text is represented based on the features later on. Each word of a textual document is considered as a feature in most cases. Still, other approaches which take syntactic and semantic rules into account for more detailed classification can be chosen (Nassirtoussi et al. 2014). After the features are selected, their dimensionality needs to be reduced, because an oversized number of features will lead to a decreased efficiency of most machine learning algorithms (Pestov 2013). Standard procedures of dimensionality reduction include stemming, conversion to lower case letters and removal of stop-words (e. g. words with less than three letters), punctuation, web page addresses and numbers (Nassirtoussi et al. 2014). In the last step of the pre-processing phase, the features need to be represented in numerical form. For this purpose, several techniques exist which are compared in Tasci and Güngör (2013). One of the most common technique is the Term Frequency-Inverse Document Frequency (TF-IDF). This value increases proportionally to appearances of a word in a document but decreases when a word occurs often in the corpus. Hence, words which are popular in general have limited relevance in text classification (Nassirtoussi et al. 2014). Results from text mining utilizing machine learning algorithms are similar to the ones attained with dictionary based approaches but are usually more accurate (Stieglitz et al. 2014). In contrast, Das and Chen (2007) analyse in how far stock prices are related to sentiment in UGC the next day. They were not able to establish a connection in this direction. Kloptchenko et al. (2004), Shirata and Sakagami (2008) and Shirata et al. (2011) analyse the textual part of annual or quarterly reports with machine learning algorithms. Kloptchenko et al. (2004) conclude that annual reports contain sentiment which predicts future financial performance. Shirata and Sakagami (2008) as well as Shirata et al. (2011) state, that it can be distinguished between bankrupt and non-bankrupt companies based on key words in previous annual reports.

It can be stated, that dictionary and machine learning based text mining has received great attention in the financial domain. Nevertheless, manual approaches and the application of text mining methods on social media data for corporate credit risk analysis have not been investigated to a great extent yet.

3 HYPOTHESES DEVELOPMENT

Concerning the amount of UGC there is theoretical support for two hypotheses. Either, the amount of UGC is greater for negative sentiment and hence for insolvent companies or it is greater for positive sentiment which is more likely to occur related to well performing companies (Anderson 1998; Bollen et al. 2011). Empirical results from Tsai et al. (2010) which are closely related to our research domain show, that firms receive worse ratings when news coverage was high. This effect is especially strong for companies with speculative ratings. Furthermore, dissatisfied customers, which communicate their experiences to a greater extent than satisfied customers, can be an antecedent for corporate insolvencies (Hall & Young 1991; Kotler 1991; Richins 1987). Therefore, we expect the amount of UGC to be greater prior to insolvencies of companies than for a reference group consisting of solvent companies in accordance with these results. Hence, we formulate our first research hypothesis as follows:

H1: The amount of UGC related to insolvent companies prior to their insolvency is greater than for solvent companies in the same time span.

Two former literature reviews on text mining in the financial domain conclude that prediction models for stock markets based on UGC can be established (Kearney & Liu 2014; Nassirtoussi et al. 2014). This should be impossible according to the efficient market hypothesis, if the information in UGC was known before social media platforms existed (Bollen et al. 2011). Then, investors would have considered it in their investment decisions (Timmermann & Granger 2004). We investigate if creditors could use this new information in order to further reduce the asymmetry of information between themselves and the debtors. Since especially negative sentiment is followed by a negative development of financial figures, sentiment from UGC ought to be worse before an insolvency when opposed to sentiment from comparable solvent companies (Chen et al. 2013; Kearney & Liu 2014; Nassirtoussi et al. 2014). We therefore formulate our second hypothesis as follows:

H2: Sentiment in UGC related to insolvent companies prior to their insolvency is worse than for solvent companies in the same time span.

In addition to that, we analyse in how far companies can be classified as prospectively solvent or insolvent based on UGC in order to establish models for decision support which can be used by creditors to further reduce the asymmetry of information between themselves and debtors. Hereby, the accuracy should exceed results from a random classification (Field 2009). In the present two category case (solvent/insolvent), a random classification would achieve an accuracy of 50%. Therefore we formulate hypothesis 3 as follows:

H3: Companies can be classified as prospectively solvent or insolvent with more than 50% accuracy based on the sentiment in UGC.

4   RESEARCH SETTING AND METHOD

In order to test the validity of the hypotheses we draw on a data set composed of insolvent and solvent German companies as well as company related UGC from the major microblogging platform Twitter. A detailed description of the data set and how it was retrieved as well as our data analysis methods from the domains of content analysis and quantitative statistics will be revealed in the following.

4.1   Data Collection

We refer to the German Federal Bureau of Statistics which discloses the ten biggest German insolvencies in the year 2013 measured by the number of affected employees for identification of insolvent companies (Federal Bureau of Statistics 2014). Then, we extracted the industries of all insolvent companies according to the official classification of industries in Germany in order to assemble an appropriate reference group (Federal Bureau of Statistics 2008). Hereby, we accessed the Hoppenstedt company database provided by Bisnode, a European vendor for economic information. After that, we selected the going-concern company in the same industry which has the most concurring number of employees and the same legal form as counterpart for every insolvent company from the database. Furthermore, we discussed industry analyses in order to ensure the comparability of each pair of companies beyond the official classification. The insolvent companies can be seen in the first column of Table 1 whereas the solvent counterparts are listed in the fourth column.

We draw on Twitter for the gathering of company related UGC since it is the most known microblogging site and it has been proven that indicators for future firm fundamentals can be extracted from the posts which are called Tweets (Bollen et al. 2011; Brown 2012; Vu et al. 2012; Zhang et al. 2012). We extracted all Tweets which are published in German one year in advance to the date on which the bankrupt company filed for insolvency for each pair of companies. We choose the time frame of one year because periodically updated information currently used in corporate credit assessment is usually updated yearly. Hence, we are able to evaluate the usefulness of UGC in order to add to the periodically updated information base. We implemented a crawler for data collection, due to limitations of the Twitter API concerning the extraction of historic Tweets. We curtailed the full name of each company insofar as it still allows a distinct matching to the company. By doing so, we are able to capture Tweets even when the full name of the company was not mentioned by the users. The complete dataset which is comprised of 7071 Tweets is summarized in alphabetical order regarding the insolvent companies in Table 1. The number in brackets represents the number of Tweets prior to the date on which the bankrupt company filed for insolvency.

Insolvent companies			Solvent Companies	
Company name	No. of Tweets	No. of Tweets		Company Name
Alpine Bau GmbH	393 (278)	101 (100)		Hochtief Solutions AG
Backstube Siebrecht GmbH & Co. OHG	30 (30)	4 (4)		Bäcker Görtz GmbH
Baumarkt Max Bahr GmbH & Co. KG	1201 (805)	139 (137)		Hagebaumarkt GmbH & Co. AG
Conergy AG	214 (183)	264 (264)		Centrotherm Photovoltaics AG
Flexstrom AG	2439 (1439)	162 (160)		Badenova AG & Co. KG
IVG Immobilien AG	555 (547)	106 (106)		PlanetHome AG
Kunert Fashion GmbH & Co. KG	6 (6)	19 (19)		Bugatti GmbH
Loewe AG	383 (352)	105 (105)		Medion AG
Praktiker Deutschland GmbH	198 (156)	216 (216)		Obi GmbH
Walter Services GmbH	367 (295)	169 (168)		Arvato direct services GmbH

Table 1.   Overview of dataset

4.2   Data Analysis

It is necessary to quantify the UGC in order to be able to adapt quantitative statistical methods and test the hypotheses (Nassirtoussi et al. 2014). We rely on the methodological approach for manual content analysis described by Neuendorf (2002) for this purpose. It is suitable to start with manual analysis of textual data because the applicability of UGC has seen little attention in the domain of credit risk analysis as pointed out earlier and manual analysis tends to be most accurate (Stieglitz et al. 2014). Furthermore, well established sentiment dictionaries are available for the English language only. In addition to that, general dictionaries for sentiment analysis cannot be applied in the financial domain without restrictions since some words have different meanings in a financial context (Li 2010).

Based on suggestions by Neuendorf (2002) we constructed a codebook with instructions for coders and the coding form in which they entered their codes. We drew on quantification techniques from Bollen et al. (2011) while creating the initial codebook. Unlike them, we did not count positive and negative tweets and calculated the ratio of both, but we assigned a weight of -1 to negative Tweets, a weight of 0 to neutral Tweets and a weight of 1 to Tweets containing positive sentiment concerning the company at the beginning. Later on, this enables us to distinguish between negative, positive and neutral sentiment in aggregated scores more clearly. A value below zero represents negative sentiment and a value above zero positive sentiment. We refined the codebook and the coding scheme in an iterative process of coder training and evaluation as recommended by Neuendorf (2002). In the final coding process, two independent coders were advised to label each Tweet with the corresponding numerical value, if the requirements listed in Table 2 applied. In contrast to the initial codebook, we decided to weight Tweets, which contain evidence concerning financial (in)stability regarding the company, with double weight towards Tweets in which the sentiment is negative/positive but financial (in)stability cannot be derived directly. This enables us to establish more differentiated sentiment scores in the following. The two coders achieved an intercoder reliability of 93,81 % which is above the level of agreement (90 %) generally accepted by researchers (Neuendorf 2002). The coders agreed on a final coding during a joint verification in which careless mistakes and different perceptions were eliminated. The following analysis is performed based on the agreed coding.

Then, we counted Tweets in each scoring category for insolvent and solvent companies based on the manual coding. Furthermore, we plotted monthly aggregated graphs for the sentiment of each company in order to gain first insights through descriptive analysis. After that, we applied t-tests in order to verify hypothesis 1 and 2. Furthermore, we made use of the k-Nearest-Neighbour algorithm in order to classify the companies based on their sentiment values as bankrupt or non-bankrupt for the verification of hypothesis 3. Hereby, we ran multiple iterations with sentiment scores from different timeframes in order to figure out in how far the classification accuracy changes when the timeframe is shifted backwards from the date of insolvency.

Requirement	Score
Tweet contains evidence concerning financial instability regarding the company	-2
Tweet does not contain evidence of financial instability but sentiment is negative regarding the company	-1
Tweet does not contain sentiment regarding the company or contains positive and negative sentiment	0
Tweet does not contain evidence of financial stability but sentiment is positive regarding the company	1
Tweet contains evidence concerning financial stability regarding the company	2

Table 2.   Codebook

5   RESULTS AND DISCUSSION

In the following, we first describe the findings of our descriptive analysis. Then, we unfold our application of t-tests and the k-Nearest-Neighbour algorithm for the verification of our hypotheses.

5.1   Descriptive Statistics

The frequency distribution of Tweets for insolvent and solvent companies in each sentiment category before the date of insolvency can be seen in Table 3. It is apparent, that the amount of Tweets is far greater for insolvent companies (4091 Tweets) than for solvent companies (1279 Tweets) which supports hypothesis 1. In average, 409.1 Tweets were posted for an insolvent company and 127.9 for a solvent company. It is remarkably, that the standard deviation of 433.5 is greater than the mean for insolvent companies. This denotes that outliers with many Tweets such as the energy provider Flexstrom AG and the home improvement market Baumarkt Max Bahr GmbH and outliers with few Tweets such as the fashion retailer Kunert Fashion GmbH and the bakery Backstube Siebrecht GmbH strongly effect the distribution. The standard deviation is 80.48 for solvent companies, which is a little greater than half of the mean. Especially companies with few Tweets such as the bakery Bäcker Görtz GmbH and the fashion retailer Bugatti GmbH influence the deviation. Presumably some branches like bakery and fashion retailers are not discussed widely in UGC. We decided not to exclude the outliers and run the quantitative tests on the whole data set due to our limited number of cases.

Furthermore, it can be comprehended, that the sentiment for insolvent companies is more negative. More than one fourth of the Tweets (26.8 %) contain evidence concerning financial instability regarding the insolvent companies. Solvent companies are characterized through neutral instead of positive sentiment since 79.4 % of Tweets related to solvent companies do not contain positive or negative sentiment at all. The mean of sentiment scores is -0.46 for insolvent companies with a standard deviation of 1.08 and 0.13 for solvent companies with a standard deviation of 0.7. The great standard deviations are caused by the interval scale which we used to measure the sentiment. They could therefore be reduced by using more fine-grained sentiment measurements.

	Sentiment Score					
	-2	-1	0	1	2	Sum
Insolvent companies	1096 (26.8 %)	355 (8.7 %)	2042 (49.9 %)	496 (12.1 %)	102 (2.5 %)	4091 (76.2 %)
Solvent companies	40 (3.1 %)	34 (2.7 %)	1015 (79.4 %)	101 (7.9 %)	89 (7.0 %)	1279 (23.8 %)

Table 3.   Frequency distribution of Tweets in sentiment categories before date of insolvency

Then, we calculated monthly sentiment scores for every company in order to visualize the development of sentiment in time. These scores are represented by the average sentiment of all Tweets within a calendar month. There are 13 monthly scores for every company because the date of insolvency was never on the last day of a month. If the date of insolvency would have been on the last day of a month, then there would be 12 monthly scores for each company in that pair only. Sentiment scores of each pair of companies are shown next to each other whereas the sentiment graph of the

insolvent company is always on the left hand side and the sentiment graph of the solvent company is on the right hand side in the following. In order to better visualize the drop of sentiment on the date of insolvency, we included Tweets on that date in the graphs.

The sentiment of the insolvent construction company Alpine Bau Deutschland AG, the energy provider Flexstrom AG and the real estate company IVG Immobilien is mostly negative rather than neutral or positive. The only positive elicitation of sentiment for an insolvent company in this group can be seen in July 2012 in the graph of Flexstrom AG. Here, a doubling of profit for the past accounting year was announced. Furthermore, the sentiment tends to decrease as time gets closer to the date of insolvency. In contrast, the scores for the according solvent companies from the reference group Hochtief Solutions AG, Badenova AG and Planet Home AG are mostly neutral with temporary positive elevations. These issues are depicted in Figures 1-6.
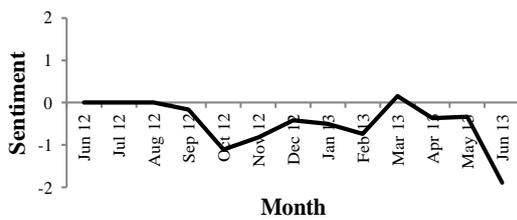
Figure 1.   *Alpine Bau Deutschland AG*
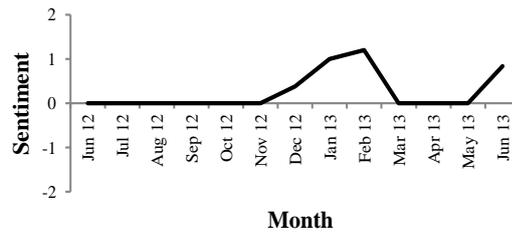
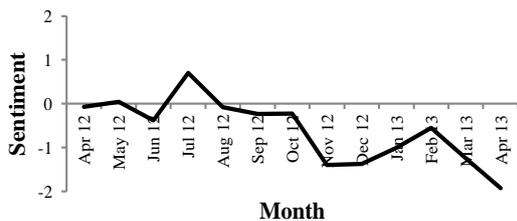*Figure 2.*   *Hochtief Solutions AG (solvent)*

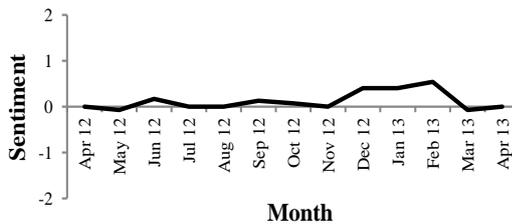*Figure 3.*   *Flexstrom AG*

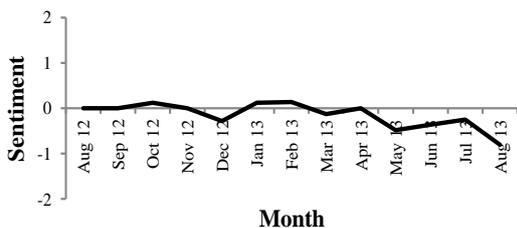*Figure 4.*   *Badenova AG*

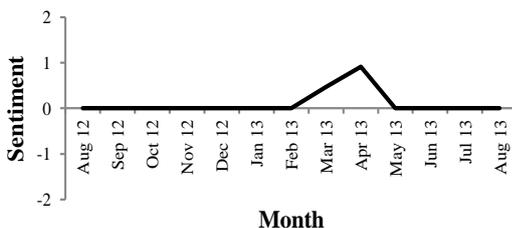
Figure 5.   *IVG Immobilien AG*

Figure 6.   *Planet Home AG*

The corresponding graphs for the companies discussed in the following paragraph can be seen in Figures 7 – 10. The depiction of sentiment scores for the insolvent home improvement market Praktiker Deutschland GmbH and the electronics producer Loewe AG show similar results as the graphs from the three insolvent companies depicted in Figures 1, 3 and 5. A cooperation with a competitor is discussed in Tweets which causes the positive peak in December 2012 and January 2013 for Praktiker Deutschland GmbH. This cooperation was expected to reduce procurement costs. Furthermore, a capital increase about 60 million euro is announced. Positive Tweets concerning Loewe include speculations about a takeover from the financially stable investor Apple and announcements from the management of Loewe which reveal, that a plan has been worked out in

order to get out of the red figures and finalize restructuring activities. Tweets related to the solvent counterparts OBI GmbH and Medion AG do not show any positive or negative sentiment at all. This indicates, that neutral sentiment can bespeak financial stability which is substantiated by findings from Bollen et al. (2011). Their results indicate, that neutral sentiment is followed by positive stock returns.
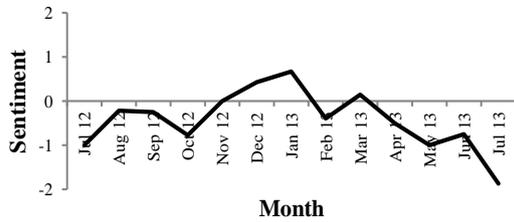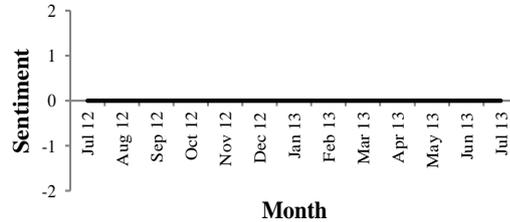

Figure 7.   Praktiker Deutschland GmbH   *Figure 8.   Obi GmbH*
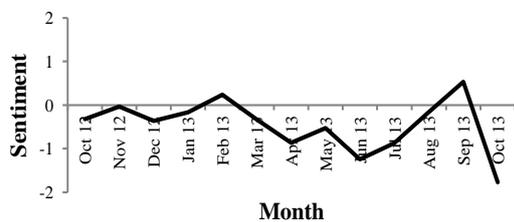

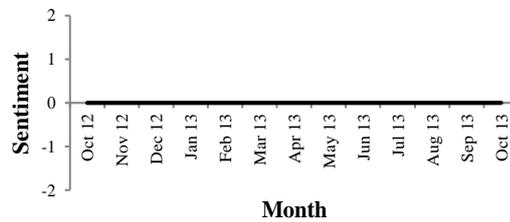
Figure 9.   Loewe AG   *Figure 10.   Medion AG*

A special case is the pair consisting of the insolvent Conergy AG and the counterpart Centrotherm Photovoltaics AG whose graphs can be seen in Figures 11 and 12. These companies are operating in the German solar industry which is subject to a major crisis (Fücks 2013). The positive Tweets shortly before the insolvency of Conergy AG are caused by announcements of capital increases, the commissioning of solar parks, and the agreement in contentions with former managers of Conergy AG. The solvent company Centrotherm Photovoltaics AG announced, that it will undergo a protective screen procedure on the 11th of July 2012 which resulted in negative sentiment. Companies can request a protective screen procedure in which they can recapitalize themselves according to German law. Therefore, a filing under regular insolvency law is circumvented. This procedure was rescinded and Centrotherm Photovoltaics was restructured successfully at the end of May 2013. Actions for restructuring were announced in August resulting in positive sentiment. Negative sentiment became prevalent again at the end of the relevant timespan. These negative Tweets are mainly caused by reports in which a customer withdraws orders. However, it can be seen, that the sentiment for the insolvent company Conergy AG is worse than for the counterpart.
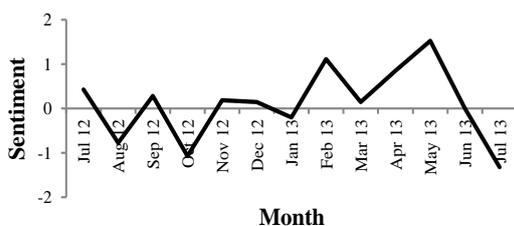

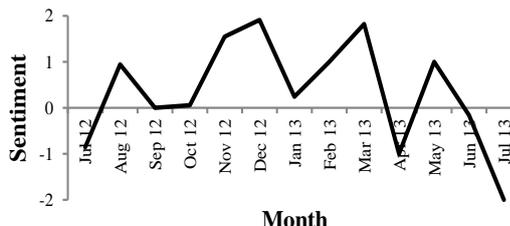
Figure 11.   Conergy AG   *Figure 12.   Centrotherm Photovoltaics AG*

We decided not to picture the graphs of the companies referred to in the following since they do not show patterns which give further insight into the issue. Instead, we want to focus on the presentation and discussion of results from the quantitative analysis instead of deepening the descriptive statistics and qualitative analysis in this paper.

All Tweets were labelled as neutral for the fashion retailer Kunert Fashion GmbH. The sentiment scores of the home improvement market Max Bahr GmbH never raise below -0.5 or above 0.5 until one month before insolvency. Here the score drops below -1.5. The same pattern is apparent for the call center operator Walter Services GmbH. The sentiment in Tweets related to the bakery Backstube Siebrecht GmbH is almost always neutral. Only in October 2012 two Tweets discuss the rental of new office area and expansion endeavours. Furthermore, in November 2012 and March 2013 one Tweet is issued in which it is stated that the company does not treat employees well.

In addition to that, the sentiment scores of the solvent counterparts for Walter Services GmbH and Backstube Siebrecht GmbH are always zero. The score of Hagebaumarkt GmbH, which is the counterparkt of Max Bahr GmbH, is zero in all but three month. In these month the score never raises above 0.5. For the Bugatti GmbH, the counterpart of Kunert Fashion Gmbh, the score reaches 1.5 in March 2012 due to announcements concerning the accomplishment of growth goals. Furthermore, in February 2013 it is stated that the success story of the company is going on, resulting in a monthly score of 1.

5.2 Quantitative Statistics

We restrained from verifying H1 and H2 with logistic regression models according to Hall and Young (1991) because the precondition of independence of errors is not fulfilled by our sample. It is required that a variable is surveyed for each case once only. Since we tested sentiment scores for multiple tweets related to each case, the results of logistic models would not be profound (Field 2009).

Instead, we tested if the mean of the number of Tweets is greater for insolvent companies than for solvent companies in order to verify the first hypothesis. Hereby, we considered Tweets before the date of insolvency only, since our goal is to anticipate insolvencies. These numbers are listed in brackets in Table 1. We applied an independent t-test for this purpose because the data is not related to dependent samples (Field 2009). A precondition of this test is a normal distribution of the data. We applied the Kolmogorov-Smirnoff test (K-S) in order to verify the assumption for our data set (Field 2009). The results of both tests can be seen in Table 4.

	Significance Level K-S	Significance Level t-test
Insolvent companies	0.071	0.072*
Solvent companies	0.200	

Table 4. Kolmogorov Smirnoff and t-test for H1

Since the significance level is above 0.05 for insolvent as well as solvent companies, the test is not significant for both variables and the null hypothesis of the K-S test, which assumes a not normal distribution, cannot be confirmed. Therefore, we further presume a normal distribution of the data which implies, that the preconditions for the t-test are fulfilled (Field 2009). The last column of Table 4 shows the result of the t-test. Hereby, it is verified whether the means are significantly different or if their difference is likely to be caused by coincidence. Established p-values need to be below 0.1, 0.05 or 0.01 in order to confirm the hypothesis (Field 2009). The p-value is 0.072 and hence, below 0.1. Therefore, we can confirm hypothesis 1 and conclude that the amount of UGC is greater for insolvent companies than for solvent companies before the date of insolvency. Hence, a rise in UGC could be a hint for financial instability and decreased credit worthiness of a company. These results are in alignment with findings from Tsai et al. (2010) who conclude that especially companies with speculative ratings and high coverage in UGC receive worse ratings in the future. Nevertheless, the hypotheses can only be confirmed on the weakest level of significance. Therefore, we recommend to verify this hypothesis with more comprehensive datasets on higher levels of significance before utilizing the result in practical applications of corporate credit risk analysis.

Hypothesis 2 is verified analogous to hypothesis 1 by an independent t-test, but instead of the amount of Tweets we compared the means of the sentiment scores in the Tweets before the date of insolvency. The central limit theorem states, that data can be explained using a normal distribution, when the quantity of observations (n) exceeds 30 (Field 2009). We tested the mean of sentiment scores for all 4091 Tweets belonging to insolvent companies against the mean of the sentiment scores of all 1279 Tweets belonging to solvent companies. Our n of 5370 exceeds 30 by far and we can assume that our data is normally distributed. The p-value of the independent t-test is below 0.001*** which indicates, that we can verify hypothesis 2 on all established levels of significance (Field 2009). This reveals, that the sentiment is worse for insolvent companies than for solvent companies. Therefore, a decrease in sentiment can be a hint for financial instability and a higher probability of payment default of a company. Many researchers have attained comparable results when matching UGC to market data such as stock prices as pointed out in the related literature section. Great amounts of negative sentiment in UGC are followed by negative returns at the stock market (Bollen et al. 2011).

We intended to apply discriminant analysis in order to verify hypothesis 3 and determine in how far companies can be classified as potentially insolvent or solvent based on the monthly sentiment scores because discriminant analysis is the dominant technique used in the domain of credit risk assessment (Grier 2007; Olson & Wu 2008). We restrained from doing so, because results from the K-S test reveal, that the precondition of normal distribution is not fulfilled by our monthly sentiment scores. Therefore, we applied the non-parametric k-Nearest-Neighbour algorithm which can also be used in order to classify companies for credit risk analysis (Olson & Wu 2008). Hereby, the algorithm is used to predict the classification of companies as solvent or insolvent based on the sentiment scores from the month prior to the insolvency till six month before insolvency. After that, we incrementally shifted the timeframe of six month one month further away from the date of insolvency. We ran the k-Nearest-Neighbour algorithm for each following time frame in order to determine in how far the classification accuracy changes, when UGC issued on earlier dates is used.

Three main parameters need to be set in order to run the k-Nearest-Neighbour algorithm. These are the metric, the number of neighbours (k) which determine the classification of new data points and at what classification accuracy the algorithm can be considered as good (Thomas et al. 2002). We used the Euclidean distance metric to measure the proximity between data points because it is used most often and has shown to be effective in the domain of credit assessment (Yu et al. 2008). We selected a feature of the program for statistical analysis IBM SPSS by which the algorithm automatically selects the best number of k. We applied a V-fold cross-validation in which the dataset is split into samples. The companies of each sample are then classified based on a training of the algorithm with all other samples. 80 % of the data is used for training and the remaining 20 % are then classified by the algorithm. Therefore, the data is split into five samples with four companies each as shown by Huang et al. (2007). The companies were randomly assigned to the samples by selecting the according checkbox in SPSS. The threshold above which a classification is considered as successful is determined by the 50 % measure stated in our third hypothesis. Hence, any result above the expected random accuracy of 50 % is good. Initial settings were maintained for all other parameters since these usually do not affect the outcome to a great extent (Thomas et al. 2002).

The results of the eight iterations are depicted in Figure 13. It is apparent, that the accuracy of the k-Nearest-Neighbour algorithm is better for sentiment scores closer to the dates of insolvency. Based on scores from the six month prior to the insolvencies, 75 % of companies can be classified as belonging to the insolvent or solvent group correctly. If sentiment scores derived from Tweets of 8 to 13 month before the insolvencies are used for classification, then the accuracy is 60 % only. It is remarkable, that the accuracy does not constantly decrease when the time window is shifted backwards from the dates of insolvency. Sentiment scores from the month 3 to 8 and 8 to 13 before insolvencies achieve a higher accuracy than scores from the preceding time windows. This indicates, that hints in UGC concerning the financial stability of companies do not continuously increase when the company is moving towards bankruptcy. Instead, there are time windows in which these hints are more widespread than in other periods. Nevertheless, a general tendency about an increase in UGC based on which it can be distinguished between solvent and insolvent companies is noticeable when the date of insolvency gets closer. The third hypotheses can be confirmed because an accuracy of more than

50 % can be achieved in all iterations. The classification accuracy is below results attained with traditional corporate credit risk analysis based on financial figures which achieve accuracies above 90 % (Danenas & Garsva 2014). Therefore, it does not seem suitable to solely rely on UGC for corporate credit risk analysis. Instead, the UGC concerning companies should be constantly monitored to reduce the information gap existing between the issuing of annual reports and company announcements. Even internal information such as previous payment behaviour is updated on an irregular basis when business transactions are finalized only. UGC provides the potential to constantly monitor debtors because it is updated in real time.
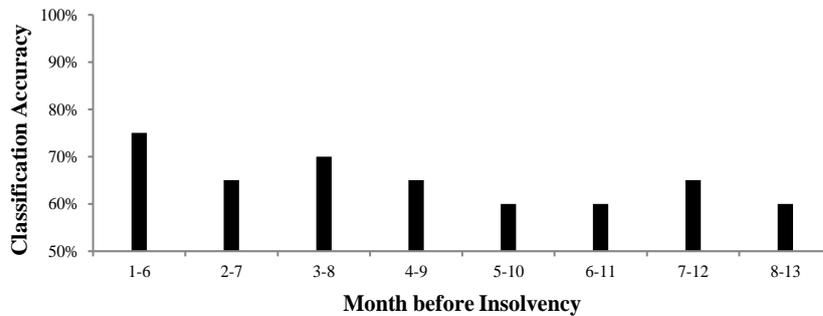

Figure 13. Classification Accuracy of the k-Nearest-Neighbour Algorithm

6 LIMITATIONS

Our results are limited by size and scope of our sample. We analysed the ten biggest insolvencies in Germany in the year 2013 and chosen counterparts only. It has to be verified if similar results can be obtained utilizing more comprehensive data sets consisting of companies with varying sizes from different industries. Furthermore, we only used data from Twitter going back one year prior to the insolvencies. UGC from other sources and longer time spans might generate different results. Especially for companies which are not covered in UGC to a great extend such as the fashion industry and bakeries, relevant data might be obtained in other social media platforms. However, analysis can only be conducted, when UGC is published for the concerning companies. In addition to that, the identification of Tweets is limited by our search terms. Users might communicate about the companies referring to colloquial names. Next to that, we restricted our study to Tweets issued in German. Analysing Tweets in other languages could lead to different or additional results. The high standard deviations indicate that our dataset is heterogeneous despite the significance of the applied tests. More homogeneous datasets should be constructed and analysed in order to further ensure the generalizability of our results. Although the coding of Tweets was verified by two coders independently, other researchers might come up with different interpretations of the textual data. Lastly, different parameter settings of the k-Nearest-Neighbour algorithm will most likely result in different classification accuracies.

7 CONCLUSION

The verification of hypothesis 1 indicates that the amount of UGC is greater for financially instable companies than for well performing companies. Furthermore, the sentiment is worse for companies which become insolvent in the future than for persisting companies because our second hypotheses is statistically verified as well. In addition to that, we are able to show that it can be distinguished between prospectively solvent and insolvent companies when the sentiment in UGC is analysed. The classification accuracy increases when the date of insolvency comes closer. Hence, UGC contains information which can be used to reduce the information asymmetry between the creditor and the debtor in a credit relationship. We contribute to the body of knowledge by applying text mining

methods to UGC in the domain of corporate credit risk analysis which has received no attention before.

The generalizability of the results should be ascertained by further studies in order to overcome the limitations by which our study is restricted. These limitations include size and scope of our sample concerning the selection of companies as well as sources and selection criterions of UGC. Furthermore it should be investigated if the classification accuracy can be improved utilizing extended data sets.

Practitioners and future research endeavours should focus on how to utilize our findings and integrate UGC as an additional source of information in state of the art corporate credit risk analysis. It is suitable to use UGC in order to constantly monitor debtors with whom ongoing credit relationships are established rather than conducting credit risk analyses solely based on UGC. This seems especially meaningful because UGC is updated in real time. In contrast to that, information used in state of the art credit assessment is updated at certain points in time only. The information gap between these points in time can therefore be reduced when companies are monitored utilizing UGC. Furthermore, it should be investigated if UGC differs between industries or business-to-business (b2b) and business-to-consumer (b2c) oriented companies. Our dataset consists of b2c companies to a great extent which are likely to have more interaction in UGC since it is mainly created by private persons (Kaplan & Haenlein 2010). Next to that, analysing different industries in our data set is not promising because there are two pairs for each industry at maximum and typical characteristics cannot be identified based on these small samples.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23 (4), 589–609.

Anderson, E. W. (1998). Customer Satisfaction and Word of Mouth. Journal of Service Research, 5 (1), 5–17.

Arrow, K. J. (1985). The Economics of Agency. In Principals and Agents: The Structure of Business (Pratt, J., Zeckhause, R. Ed.), pp. 37–51, Harvard Business School Press, Boston.

Bartels, J. C. (2013). Credit Scoring in the Cross-hairs of Consumerism. Password, 28 (4), 9.

Beck, T. R. (2014). Inkassounternehmen und der Erfolg beim Forderungseinzug. Springer Gabler, Wiesbaden.

Bollen, J., Mao, H., and Zeng, X.-J. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2 (1), 1–8.

Brown, E. D. (2012). Will Twitter Make You a Better Investor? A Look at Sentiment, User Reputation and Their Effect on the Stock Market. In Proceedings of the 15th Southern Association of Information Systems SAIS (G. Schell, A. Gardiner, & A. Ciganek Eds.), pp. 36–42, Atlanta.

Chee, N. K., Smith, J. K., and Smith, R. L. (1999). Evidence on the Determinants of Credit Terms Used in Interfirm Trade. The Journal of Finance, (3), 1109–1129.

Chen, H., De, P., Hu, Y., & Hwang, B. (2013). Customers as advisors: The role of social media in financial markets. Retrieved April 07, 2014, from http://web.ics.purdue.edu/~bhwang/customers.pdf.

Danenas, P., and Garsva, G. (2014). Selection of Support Vector Machines based classifiers for credit risk domain. Expert Systems with Applications, 42 (6), 3194–3204.

Das, S. R., and Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science, 53 (9), 1375–1388.

Federal Bureau of Statistics (2008). Classification of industries. Retrieved December 12, 2014, from https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content7 5/KlassifikationWZ08.html.

Federal Bureau of Statistics (2014). Die zehn größten Unternehmensinsolvenzen in Deutschland im Jahr 2013. Retrieved September 11, 2014, from http://de.statista.com/statistik/daten/studie/12997/umfrage/groesste-unternehmensinsolvenzen-in-deutschland/.

Field, A. (2009). Discovering Statistics using SPSS. 3rd Edition. Sage Publications, London.

Fücks, R. (2013). Intelligent wachsen. Carl Hanser, München.

Galindo, J., and Tamayo, P. (2000). Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications. Computational Economics, 15 (1-2), 107–143.

Graham, A., and Coyle, B. (2000). Corporate Credit Analysis. Fitzrov Dearborn Publishers, Chicago.

Grier, W. A. (2007). Credit Analysis of Financial Institutions. 2nd Edition. Euromoney, London.

Hall, G., and Young, B. (1991). Factors Associated with Insolvency Amongst Small Firms. International Small Business Journal, 9 (2), 54–63.

Henry, E. & Leone, A. J. (2009). Measuring Qualitative Information in Capital Markets Research. Retrieved July 07, 2014, from http://ssrn.com/abstract=1470807.

Hirokawa, S., Baba, T., and Nakatoh, T. (2011). Search and Analysis of Bankruptcy Cause by Classification Network. Model and Data Engineering, (6918), 152–161.

Huang, C.-L., Chen, M.-C., and Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications, 33 (4), 847–856.

Jegadeesh, N., and Di Wu (2013). Word power: A new approach for content analysis. Journal of Financial Economics, 110 (3), 712–729.

Kabbage (2015). How Kabbage Loans Work. Retrieved January 19, 2015, from https://www.kabbage.com/how-it-works/.

Kaplan, A. M., and Haenlein, M. (2010). Users of the world unite! The challenges and opportunities of Social Media. Business Horizons, 53 (1), 59–68.

Kearney, C., and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171–185.

Keßler, R. (2010). Einführung des Kundenratings im Kreditmanagement eines Handelsunternehmens. In Debitorenrating (G. S. Becker & O. Everling Eds.), pp. 109–123, Gabler, Wiesbaden.

Kloptchenko, A., Eklund, T., Back, B., Karlsson, J., and Vanharante, H. (2004). Combining Data and Text Mining Techniques For Analyzing Financial Reports. Intelligent Systems in Accounting, Finance and Management, 12 (1), 29–41.

Kotler, P. (1991). Marketing Management: Analysis, Planning, Implementation, and Control. Prentice Hall, Sydney.

Kramer, J. W., and Nitsch, K. W. (2010). Facetten der Unternehmensfinanzierung. CT Salzwasser-Verlag GmbH, Bremen.

Li, F. (2010). The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach. Journal of Accounting Research, 48 (5), 1049–1102.

Nassirtoussi, A. K., Wah, T. Y., Aghabozorgi, S. R., and Ling, D. N C. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41 (16), 7653–7670.

Neuendorf, K. A. (2002). The Content Analysis Guidebook. Sage Publications, Chicago.

Olson, D. L., and Wu, D. (2008). New Frontiers in Enterprise Risk Management. Springer, Heidelberg.

Pestov, V. (2013). Is the -NN classifier in high dimensions affected by the curse of dimensionality? Computers and Methematics with Applications, 65, 1427–1437.

Richins, M. L. (1987). A multivariate analysis of responses to dissatisfaction. Journal of the Academy of Marketing Science, 15 (3), 24–31.

Schumann, M. (2002). Bonitätsbeurteilungen und Kreditprüfungen im Electronic Business. In Electronic Business: Theoretische Aspekte und Anwendungen in der betrieblichen Praxis (R. Gabriel & U. Hoppe Eds.), pp. 413–429, Physica-Verlag, Berlin.

Shirata, C. Y., and Sakagami, M. (2008). An Analysis of the "Going Concern Assumption". Journal of Emerging Technologies in Accounting, 5, 1–16.

Shirata, C. Y., Takeuchi, H., Ogina, S., and Watanabe, H. (2011). Extracting Key Phrases as Predictors of Corporate Bankruptcy. Journal of Emerging Technologies in Accounting, 8, 31–44.

Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. (2014). Social Media Analytics. Business & Information Systems Engineering, 6 (2), 89–96.

Stiglitz, J. E. (1988). Principal and Agent. Woodrow Wilson School of Public and International Affairs, Princeton University, New Jersey.

Stiglitz, J. E., and Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. The American Economic Review, 71 (3), 393–410.

Tasci, S., and Güngör, T. (2013). Comparison of text feature selection policies and using an adaptive framework. Expert Systems with Applications, 40, 4871–4886.

Thomas, L. C., Edelman, D. B., and Crook, J. N. (2002). Credit Scoring and Its Applications. Society for Industrial and Applied Mathematics, Philadelphia.

Timmermann, A., and Granger, C. (2004). Efficient market hypothesis and forecasting. International Journal of forecasting, (20), 15–27.

Tsai, F.-T., Lu, H.-M., and Hung, M.-W. (2010). The Effects of News Sentiment and Coverage on Credit Rating Analysis. In Proceedings of the 14th Pacific Asia Conference on Information Systems (PACIS), pp. 1963–1970, Taipei.

Vu, T. T., Chang, S., Ha, Q. T., and Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in twitter. In Workshop on Information extraction and Entity Analytics on Social Media Data, pp. 23–38, The COLING 2012 Organizin Committee, Mumbai.

Yu, L., Wang, S., Lai, K. K., and Zhou, L. (2008). Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machines. Springer Verlag, Heidelberg.

Zhang, X., Fuehres, H., and Gloor, P. (2012). Predicting Asset Value through Twitter Buzz. In Advances in Collective Intelligence (J. Altmann, U. Baumöl, & B. J. Krämer Eds.), pp. 23-34, Springer, Berlin.