The Financial Credit Risk Assessment Model: Three Perspectives

Eric Mantelaers
Optimizing Knowledge-Intensive Business Processes
Zuyd University of Applied Sciences
Sittard, the Netherlands
eric.mantelaers@zuyd.nl

Martijn Zoet
Optimizing Knowledge-Intensive Business Processes
Zuyd University of Applied Sciences
Sittard, the Netherlands
martijn.zoet@zuyd.nl

Abstract—**Within recent years, Financial Credit Risk Assessment (FCRA) has become an increasingly important issue within the financial industry. Therefore, the search for features that can predict the credit risk of an organization has increased. Using multiple statistical techniques, a variance of features has been proposed. Applying a structured literature review, 258 papers have been selected. From the selected papers, 835 features have been identified. The features have been analyzed with respect to the type of feature, the information sources needed and the type of organization that applies the features. Based on the results of the analysis, the features have been plotted in the FCRA Model. The results show that most features focus on hard information from a transactional source, based on official information with a high latency. In this paper, we re-address and -present our earlier work [1]. We extended the previous research with more detailed descriptions of the related literature, findings, and results, which provides a grounded basis from which further research on FCRA can be conducted.**

Keywords-Financial Credit Risk Assessment; Business Failure Prediction; Credit Risk Features; DMN Requirements Diagrams (DRD).

I. INTRODUCTION

Within the field of the Financial Credit Risk Assessment (FCRA) there are two main areas of interest. The first main interest, credit rating (or scoring), is used to solve the problem to label companies as bad/good credit or bankrupt/healthy. Credit rating is used not only internally for screening borrowers, pricing loans and managing credit risk thereafter, but also externally for calibrating regulatory capital requirements [2]. The second main interest, bankruptcy (failure) prediction (or business failure prediction or going concern assessment) is intended to predict the probability that the company may belong to a high-risk group or may become bankrupt during the following year(s). Both of them are strongly related and solved in a similar way, namely as a binary classification task. In this paper, both categories of problems are collectively called FCRA, which is a business decision-making problem that is relevant for creditors, auditors, senior management, bankers and other stakeholders.

FCRA is a domain which has been studied for many decades. According to Balcaen and Ooghe [3], there are four main areas with reference to FCRA: (1) Classical paradigm (arbitrary definition of failure, non-stationarity and data instability, sampling selectivity), (2) Neglect of the time dimension of failure (use of one single observation, fixed score output/concept of resemblance/descriptive nature, failure not seen as a process), (3) Application focus (variable selection, selection of modelling method), (4) Other problems (use of a linear classification rule, use of annual account information, neglect of multidimensional nature of failure). The literature on FCRA and business failure dates back to the 1930's [27]. Watson and Everett [4] described five categories to define failure: 1) ceasing to exist (discontinuance for any reason), 2) closing or a change in ownership, 3) filing for bankruptcy, 4) closing to limit losses and 5) failing to reach financial goals. When the FCRA is negative, it is called business failure, which is a general term and, according to a widespread definition, it is the situation that a firm cannot pay lenders, preferred stock shareholders, suppliers, etc., or a bill is overdrawn, or the firm is bankrupt according to the law [5]. There is extensive literature in which this topic has been researched from the perspective of auditors or bankers. On the other hand, rare literature can be found about related literature from an information and decision perspective. The features (variables) which are relevant in the field of FCRA will be analyzed in this paper. In this paper the focus will be on the auditor's, bankers and crediting rating firms, hence forward the term financial industry will be used to describe all three. A combination will be made between the financial industry and an information and decision perspective.

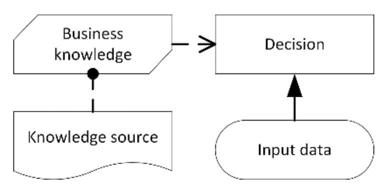

Figure 1. DRD-level Elements

To do so, the DRD model will be used. The reason DMN (Decision Model and Notation) is used, is because it is currently the standard to model decisions. In September 2015, the Object Management Group (OMG) [6] released a new standard for modelling decisions and underlying business logic, DMN The DMN standard is based on two levels; the Decision Requirements Diagram (DRD) level and the Decision Logic Level (DLL). The DRD level consists of four concepts that are used to capture essential information with regards to decisions: 1) the decision, 2) business knowledge, which represents the collection of business logic required to execute the decision, 3) input data, and 4) a knowledge source, which enforces how the decision should be taken by influencing the underlying business logic. The contents of the DLL are represented by the business knowledge container in the DRD level.

The remainder of this paper is organized as follows. Section II contains a description of relevant literature regarding features, feature selection, and techniques with reference to FCRA, from a combined perspective of both the financial industry and information and decision analysts, followed by the research method in Section III. In Section IV, our data collection and analysis will be reported. Subsequently, in Section V, a presentation of the results derived from the applied data analysis techniques will be given. The conclusion (Section VI) closes the article.

II. LITERATURE REVIEW

Feature selection is a critical step in FCRA, which refers to the process that reduces the feature space and selects an optimum subset of relevant features. Three possible methods can be distinguished: 1) human, 2) statistical and 3) hybrid. In the human approach, an auditor decides which features are important and how they relate to each other. The model in 'the head' of the auditor is rebuild into the system. For the statistical approach several alternative methodologies are applied for the feature selection. Tsai [7] compares five well-known feature selection methods used in bankruptcy prediction, which are: 1) *t*-test, 2) correlation matrix, 3) stepwise regression, 4) principle component analysis (PCA) and 5) factor analysis. The hybrid approach applies both the human and statistical manner.

Statistical techniques:
1. Linear discriminant analysis (LDA)
2. Multivariate discriminate analysis (MDA)
3. Quadratic discriminant analysis (QDA)
4. Logistic regression (LR)
5. Factor analysis (FA)

Intelligent techniques:
1. Neural networks (NN)
2. Decision trees (DT)
3. Rough sets
4. Case-based reasoning (CBR)
5. Support vector machines (SVM)
6. Data envelopments analysis
7. Soft computing (hybrid intelligent systems)
8. Operational research techniques
9. Other intelligent techniques

Figure 2. Statistical and Intelligent Techniques

To apply the selected features from the features selection to take the FCRA-decision, different methods are applied. Broadly, these methods are dived into two broad categories: statistical and intelligent techniques [8] [9]. They exist out of multiple sub-categories, see Figure 2. For a detailed description of the techniques we refer to Ravi Kumar and Ravi [8].

Based on literature studied, we developed a model that exists out of three axes that determine the type of features applied. To ground our theory, we first present the end model: the FCRA Model.

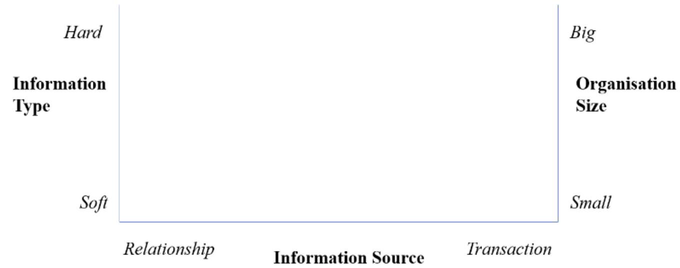

Figure 3. Financial Credit Risk Assessment Model

The first axe describes the type of data that organizations retrieve to make a judgement about the financial credit risk. In the papers of Berger [21][25], the same distinction is made in an information type perspective: hard versus soft (or quantitative versus qualitative) data. Different related names are used in this field:

Quantitative features	Qualitative features
Hard information	Soft information
Financial information	Non-financial information
Accounting information	Non-accounting information

The second axe describes the manner in which this information is retrieved. For example, two manners in which information can be collected are: 1) through face to face contact between a loan officer and the organization's owner and 2) through a form on a website or any other digital manner. The third axe describes the organization size, varying from small to big. The loan decision model of small banks is known to differ from the loan decision model of large banks [12]. According to Berger [10] small organizations (organization size), make use of soft information (information type), based on the relationship with their clients (information source). Bollen et al. [13] recognize four categories of business failures: 1) Tadpole (company failed because it was a basically unhealthy company, 2) Drowned frog (the company is over-ambitious or may show signs of extreme growth, 3) Boiled frog (companies in this category may be failing as a result of external conditions (e.g., disasters), bad economic conditions or fundamental changes in the business environment to which the company has failed to respond adequately, 4) Bullfrog (the companies in this category have drawn a relatively large portion of public attention, because they are often related to fraudulent activities of managers or employees).

A. *Information Type*

Hard

According to Petersen [14] hard information is almost always recorded as numbers and is comparable. The durability of information is potentially greater when it is hard. The collection method of hard information is mostly not personal. Hard information is mostly standardized and easy to document and transfer to others [15]. Nemoto et al. [16] also recognize the verifiability which normally is higher in case of hard information. Decision processes which depend upon hard information are easier to automate. Knowing what information you are looking for, and why it is valuable, is essential if information collection and possibly decision making, based on the information is to be delegated. Most features are based on data from the financial statements. Financial statements are, in most organizations, created once or twice a year. Therefore, the data needed to calculate the features is available once or twice a year. This causes an information opacity problem, thereby reducing the effectiveness of the features. Other organizations that also assess the financial credit risk of an organization are banks, credit assessors, etc. Both previously also had to trust numbers that are published once a year. Since this time period is too long for both parties, they searched for solutions to address this problem. According to Berger and Udell [17] hard information may include, as examples, financial ratios calculated from audited financial statements; credit scores assembled from data on the payment histories of the small and medium sized entities (SME) end its owner provided by credit bureaus; or information about accounts receivable from transparent, low-risk obligors that may pledged as collateral by the SME or sold to the financial institution.

Soft

Soft information is mostly relationship-based and not easily quantified [18]. The replacement of soft with hard information inevitably results in a loss of information. The early studies for FCRA were univariate (a specific statistical method applied) studies which had important implications for future model development. These laid the groundwork for multivariate studies. Ravi Kumar and Ravi [8] identify statistical and intelligent techniques to solve the bankruptcy prediction problem. For each type of technique, they describe the way they work. Chen, Ribeiro and Chen [9] summarize the traditional statistical models and state-of-the-art intelligent methods. Auditors can utilize data mining techniques to analyze external (soft) data (e.g., census data, social media, news articles) in their assessments of client business risk, fraud risk, internal controls, going concern [19]. Lu et al. [20] explain the possibilities of data mining (text mining) based on soft information on websites and in financial reports to predict bankruptcy.

Altman et al. [21] describe the value of qualitative (soft) information in SME risk management. They find that qualitative data relating to such variables as legal action by creditors to recover unpaid debts, company filing histories, comprehensive audit report/opinion data and firm specific characteristics, make a significant contribution to increasing the default prediction power of risk models, built specifically for SMEs. Lenders must invest in the production of ‘soft information’ to supplement the financial data used in these models [22]. Dainelli et al. [23] give a summarization of determinants of SME credit worthiness under Basel rules. As their model does not include qualitative information, future research could aim to set out the qualitative determinants in the rating judgment. Petersen [14] concludes that technology is changing the way we communicate. One of these changes is a greater reliance on hard relative to soft information. Despite this, very little research has been published on the concept of activities used by lenders to gather soft information [24]. Suter [24] studied the collection of soft information by small community banks. He built a conceptual framework existing of four factors to reduce the asymmetric information: 1) Knowledge of business, 2) Knowledge of industry, 3) Knowledge of local market, and 4) Value of the social contract. Angilella and Mazzù [25] structured the non-financial criteria hierarchically on the basis of the risk areas, specific to an innovative firm: development, technological, market, and production. The risk areas considered are: Technological risk, Market risk, Production risk, Innovation indicators, Financial criteria.

Performance

To measure performance, there are several metrics [9]. One of the most important measures is accuracy. In terms of performance, an accuracy rate between 81 and 90% reflects a realistic average performance based on the results of the analyzed studies [26]. The top five bankruptcy models with an accuracy level of more than 80 per cent are [27]: 1) Altman [28], 2) Edmister [29], 3) Deakin [30], 4) Springate, [28] and 5) Fulmer [29]. All of these only use hard features. Chen et al. [31] find that the use of soft information significantly improves the power of default prediction models.

The same conclusion is realized by Ju and Sohn [32] who proposed to update the credit scoring model based on new features like management, technology, marketability, and business and profitability. Kosmidis and Stavropoulos [33] even got one step further in their conclusion, as they state that factors such as economic cycle phase, cash flow information and the detection of fraudulent financial reporting can evidently enhance the predictive power of existing models. Altman, Sabato and Wilson [21] reach the same conclusion as they state: *“that qualitative data relating to such variables as legal action by creditors to recover unpaid debts, company filing histories, comprehensive audit report/opinion data and firm specific characters make a significant contribution to increasing the default prediction power of risk models built specifically for SMEs.”*. This leads us to the first conclusion that the financial industry should not only rely on hard features, which have a time delay, but also on soft information to assess the financial credit risk; see bottom left side in Figure 3. Relationship lending is based on soft information and is best suited for entities that are more opaque; and transactions-based lending is best suited for SMEs that are more transparent [34].

To realize proper research in this area, the researchers have to go beyond the already cumulative features and look at the base data. E.g., no longer apply the cumulative feature: current assets but instead build features on the base information such as debtors' information.

B. Information Source

In addition to the type of information available, the data source and its fluidity are also factors. In financial literature, this phenomenon is called "the hardening of soft information" [45]. The concept "the hardening of soft information" states that because personal contact with financial institutions has decreased, therefore they rely more and more on hard quantitative information. Since more banks, credit organizations, and accountants rely on non-personal contacts, this statement is gaining importance.

Thereby underlying the fact that the traditional features are the most useful features to analyze the financial credit risk. The main reason they state to support their claim is the adoption rate of technology.

However, a counter claim can be made that through the adoption of technology, soft information can be more easily collected. For example, through firehose access to social media websites. However, this will depend on the type of soft or hard information one wants to retrieve because not all soft information can be retrieved through social websites, some still might need to be retrieved face to face. Therefore, the bottom part of our model, see Figure 3, indicates the lending technologies, being the manner in which the information is retrieved. A lending technology is *"a set of screening and underwriting policies and procedures, a loan contract structure, and monitoring strategies and mechanisms"* [11]. Examples of lending technologies they apply are: leasing, commercial real estate lending, residential real estate lending, motor vehicle lending, and equipment lending, asset-based lending, financial statement lending, small business credit scoring, relationship lending and judgment lending.

C. Organisation Size

In FCRA literature, from a banking perspective, a distinction is made between the manner in which small and big banks assess the risk. Small banks apply more of a relationship perspective to assess the risk, while big banks apply the analysis of transactions to determine the risk. Although this specific distinction cannot be found in accountancy and lending (firms) literature, the hypothesis is that the same basic rules apply. Therefore, the right axe of the FCRA Model contains the size of the firms, assessing the risk; see Figure 3.

Loans to small businesses have traditionally been based on intimate relationships between borrower firms and lenders, because many of these firms are much more informationally opaque than large firms. Thus, lenders primarily rely on soft information, gathering through long-lasting transaction relationships. For banks it is difficult to obtain detailed information from small firms since the financial reports of small firms are mainly for tax purposes [35].

III. Research Method

The goal of this research is to identify and classify features that have been applied to assess Financial Credit Risk. In addition to the goal of the research, also, the maturity of the research field is a factor in determining the appropriate research method and technique. Based on the number of publications and identified features, the maturity of the FCRA research field can be classified as mature. Mature research fields should A) focus on further external validity and generalizability of the phenomena studied, or B) focus on a different perspective on the constructs and relationships between identified constructs [36]. Current studies have focused on two elements: 1) selecting the best features to predict bankruptcy, while other studies have focused on 2) comparing the efficiency and effectiveness of the different features identified. However, current analysis focuses on two viewpoints: 1) a high abstraction level and 2) a high latency perspective.

Summarized, to accomplish our research goal, a research approach is needed in which the current features are explored, compared and mapped to the FCRA Model.

To accomplish this goal, a research approach is needed that can 1) identify features for FCRA, 2) identify similarities and dissimilarities between features, and 3) map the features to the FCRA Model. The first two goals are realized by applying a structured literature research and the use of a comparison table. The last goal is realized by coding the features identified, based on a priori coding scheme.

IV. Data Collection and Analysis

As stated in the previous section, the goal of this research is to 1) identify features for FCRA, 2) identify similarities and dissimilarities between features for FCRA, and 3) map the features to the FCRA Model.

The selection of the papers has been conducted via the link-tracing methodology [37], more specifically via snowball sampling. The snowballing was applied to take advantage of the social networks of identified respondents to provide a researcher with an ever-expanding set of potential contacts [38]. Snowballing is an effective and efficient form of contact tracing for use in diversity of research methods and designs, and apparently well suited for a number of research purposes [39] - [40]. For both the hard features and soft features two different snowball samplings have been conducted. For the hard features this resulted in 238 papers that were included in the coding. With respect to the soft features this resulted in 20 papers to be selected for coding. For a study to be selected for coding, the study must explicitly address hard and/or soft features for FCRA (see Table II for details). The unit of analysis for coding is a single feature, implying that one study can contribute multiple units of analysis. For example, Alam et al. (2000) contributed five features: 1) "*Net loan losses / Total assets less Total loans*", 2) "*Net loan losses / Total loans*", 3) "*Net loan losses plus Provision for loan losses / Net income*", 4) "*Loans past due 90 days plus Nonaccrual loans / Total assets*" and, 5) "*Net income / Total assets*". This resulted in the identification of 700 hard features and 135 soft features. Each of the hard and soft features have been added to a comparison table, see Table I [2][41][42][43].

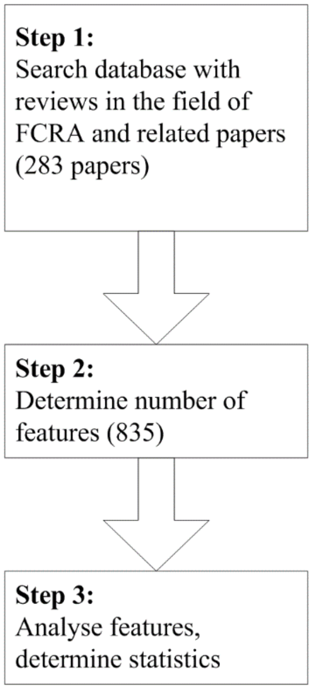

Figure 4. Feature Selection

Data analysis was conducted in one cycle of coding with the use of a priori coding scheme. The reason an a priori coding scheme was applied, is because the concepts that needed to be coded were known upfront, based on the previously defined FCRA Model.

To code the selected items, the following questions are asked: 1) *is the feature a hard or soft feature?* and 2) *is the feature a relational or transactional feature?* For example, the feature *"net income/total assets"* is a hard feature from a transactional perspective. A hard feature because the ratio can be calculated and transactional, because the figures can be derived from a system. An example of a hard / relational feature is *"the number of times the annual financial statements are deposited too late"*. A hard feature because the number can be calculated and relationship because it's a proxy of a soft feature, for example of management quality.

"The quality of management" is a soft feature from a relational perspective. A soft feature because it cannot be calculated directly and the (qualitative) information has to gathered via personal contacts.

TABLE I. SNAPSHOT COMPARISON TABLE

	Grunert (2005)	Altman (2010)	Cornee (2014)	Tronnberg (2014)
Management Quality	1		1	1
Country Court Decisions		1		
Audit Accounts		1		
Late Filling Days		1		
Sector				
Industry Risk				

V. RESULTS

In this section, the results of the data collection are presented. As described in the previous section, first features from existing studies have been collected, added to a comparison table and coded. Therefore, three separate results can be identified: 1) descriptive statistics for hard features, 2) descriptive statistics for soft features and, 3) the mapping of the hard and soft features to the FCRA Model.

A. Results from an information type perspective

As stated in this section, most features are based on data from the financial statements. Financial statements are, in most organizations, created once or twice a year. Therefore, the data needed to calculate the features, is available once or twice a year. This causes an information opacity problem thereby reducing the effectiveness of the features. Other organizations that also assess the financial credit risk of an organization are banks, credit assessors, etc. Both previously also had to trust numbers that are published once a year. Since this time period is too long for both parties, they searched for solutions to address this problem.

Camaco-Miñano et al. [44] show that sector, size, number of shareholdings, ROA, and liquidity can explain the bankruptcy process outcome and also predict the process for still-healthy firms. Three of five features exist of qualitative information.

A1. Descriptive statistics for hard features

The extraction of the features resulted in the registration of 700 features from 238 papers. From this sample, the top ten features were identified and selected; see Table II.

TABLE II. TOP TEN FEATURES

Feature 01: Net income/total assets	85 (papers)
Feature 02: current ratio	74
Feature 03: EBIT/total assets (*)	65
Feature 04: retained earnings/total assets (*)	62
Feature 05: working capital/total assets (*)	60
Feature 06: sales/total assets (*)	46
Feature 07: quick ratio	41
Feature 08: current assets/total assets	39
Feature 09: total debt/total assets	39
Feature 10: cash/total assets	32

Analysis of the hard features show that each of the top ten applied features are features that are applied in the financial statements of the organization. In addition, each feature lies a connection between the three main components of the financial statements namely: the cash flow statement, profit and loss statement, and the balance sheet. They do so by comparing the liquidity (features 02, 05, 07, 08, 09 and 10), the solvency (features 03 and 04) and the profitability (features 01, 03 and 06). Where the liquidity is primarily related to cash flow; the solvency is related to the balance sheet; the profitability is primarily related to the profit and loss statement. Of course, there is a main interrelationship between all these three main components of the financial statements.

Additionally, results show that our findings are in line with statements made in previous research, namely that the Altman model for bankruptcy prediction [28] is the most applied one. This is indicated by the fact that 4 features in the top 10 (indicated by an asterisk) are part of the Altman-Z score. And the fifth feature by Altman (Market Value of Equity/Total Liabilities) ranks thirteenth.

A2. Descriptive statistics for soft features

The extraction of the features resulted in the registration of 135 features from 20 papers. Likewise, to the hard features a top ten can be derived. However, in contrast to the hard features this top ten would exist out of features that are only mentioned four, three, or two times. One feature is listed four times, namely *"management quality"*. Four features are listed three times, namely: *"county court judgements"*, *"decision to check audited accounts"*, *"decision to issue cash flow statements"* and, *"late filing days"*. Followed by twelve features mentioned only twice. The remaining 118 features all are mentioned once. Therefore, creating a top ten didn't seem useful. In addition, the soft features have been additional coded to create a categorization (see Figure 5). The categorization has three main differentiations: 1) internal, 2) external and 3) social contract. 'Internal' is defined as qualitative (soft) information about the client; for example, about the client's management and its innovative power. 'External' can be seen as the environment that affects and interacts with the client. There are three main sub-categories: business (e.g., the number of visits with customer vendors & suppliers or visits with customers about business status), industry (e.g., the number of reviews of trade journals from customer's industries or the number of memberships in trade associations relating to customer's industries) and economy (e.g., the number of attendances at local chamber events or number of memberships in civic and community organizations). The last differentiation is 'social contract', which is defined as qualitative (soft) information about the lending relationship. 'Social contract' is further divided in two categories: quality of the credit relationship and value of the social contract.

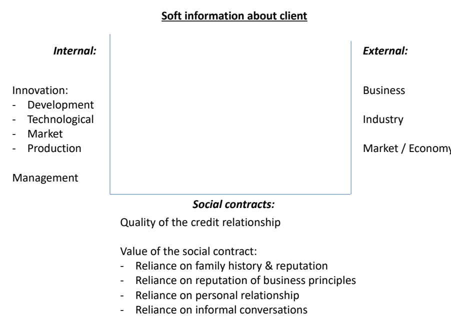

Figure 5. Categorization Soft Features

A3. Descriptive statistics for techniques

Based on the 128 papers [8] the frequency of the techniques applied to take the FCRA-decision have collected. In total 84 techniques have been identified. Seven techniques occur more than five times. Out of these seven techniques four occur more than 10 times, see Table III. The remaining 77 techniques occur up to 4, 3, 2 or 1 times.

TABLE III. TOP 7 TECHNIQUES

Technique	#
BPNN (Back propagation trained Neural Network)	28
DA (Discriminant Analysis)	18
LOGIT	18
LDA (Linear Discriminant Analysis)	10
Rough Set	08
GA (Genetic Algorithm)	05
Probit	05

A4. Descriptive Statistics for performance

Overall can be stated that research on feature identification does not clearly report on the (overall) performance of the features identified. Off the researched paper, only 16 report extensively on the performance of the applied techniques. We argue that further research should report on the performance of the identified and tested features. To measure the performance Chen [9] identified 17 measurements which can be applied, for example: 1) accuracy, 2) root mean squared error, 3) true positive, and 4) true negative. For the detailed description of these 17 measurement we refer to Chen [45].

B. Results from a DMN perspective

Analyzing the top ten features from a DMN perspective shows four results. The first result: decision versus data input show that each feature is treated like a decision. The feature is derived from one or more conditions. For example, the first feature is derived out of two conditions: net income and total assets to which a mathematical formula is applied, in this specific case, net income divided by total assets. Each feature in the 10 retrieves the applied conditions from one data source, namely, the financial statements (the cashflow statement, the profit and loss account and/or the balance sheet).

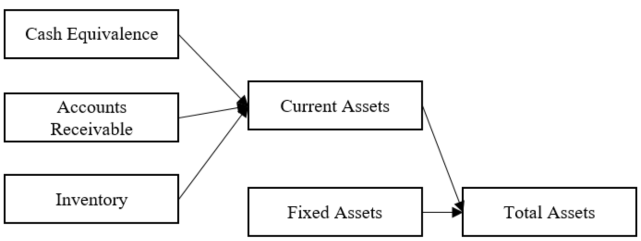

Figure 6. DRD-level Elements (Hard Features)

From the perspective of the financial statements, the conditions applied, e.g., net income, actually are data input since all are listed there. However, when analyzing one step deeper, each data input on the balance sheet or the profit and loss account is actually a decision. For example, total assets, is calculated as current assets plus fixed assets; see Figure 6. When analyzing all of the quantitative features selected, all features are derived from the cashflow statement, the profit and loss account and/or the balance sheet. A potential explanation of this phenomenon can be that the financial industry only looks at formal documents and formal statements. However, this raises the question if these combined features contain specific sub-decisions or specific input data elements that make them suitable for analysis. According to the researchers, this would be a subject to further investigate.

In addition, the features only apply information from the current financial statements. Formally, the cashflow statement, the profit and loss account and the balance sheet have to be created once a year. Most companies create this information more times a year, voluntarily or obligatory. Also, not comparing information from early years, thereby indicating that the patterns have no additional information value. By analyzing the deeper layers underneath the features described previously, the hypothesis is that a better and quicker FCRA can be performed.

C. Results from an information source perspective

The third perspective from which factors can be classified is the information source perspective. The concept *"the hardening of information"* states that because personal contact with the bank has decreased the banks rely more and more on hard quantitative information. However, if the model on which they base these conclusions is further dissected, two axes can be distinguished: A) the type of data and B) the manner in which the data is retrieved. The first axe describes the type of data that organizations retrieve to make a judgement about the financial credit risk. In the papers of Berger [21][25], the same distinction is made in an information type perspective: hard versus soft data. The second axe described the manner in which this information is retrieved. For example, two manners in which information can be collected are: 1) through face to face contact between a loan officers and the organization's owner and 2) through a form on a website or any other digital manner. Since more banks, credit organizations, and accountants rely on the second, the statement of "the hardening of information" is that only quantitative data is used. Thereby underlying the fact that the traditional features are the most useful features to analyze going concern assessment. The main reason they state to support their claim is the adoption rate of technology.

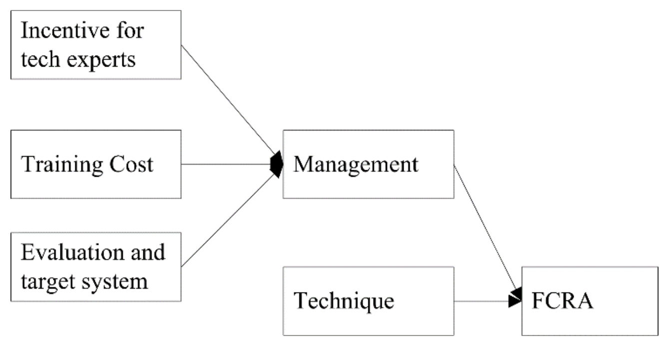

Figure 7. DRD-level Elements (Soft Features)

However, a counter claim can be made that through the adoption of technology soft information can be more easily collected. For example, through firehose access to social media websites. However, this will depend on the type of soft or hard information one wants to retrieve because not all soft information can be retrieved through social websites, some still might need to be retrieved face to face. Therefore, the bottom part of our model, see Figure 3, indicates the manner in which the information is retrieved.

D. Results from an organization perspective

In FCRA literature, from a banking perspective, a distinction is made between the manner in which small and big banks assess the risk. Small banks apply more of a relationship perspective to assess the risk while big banks apply the analysis of transactions to determine the risk. Although this specific distinction cannot be found in accountancy and lending (firms) literature, the hypothesis is that the same basic rules apply. Therefore, the right axe of the FCRA Model contains the size of the firms assessing the risk; see Figure 3.

VI. Conclusion and Future Work

In this paper, we aimed at finding an answer to the following research question: *"how to categorize financial credit risk features such that an integrative relationship is established with the information type applied and information sources used?"* To accomplish this goal, we conducted a literature study to identify features that have been designed and applied in previous research followed by coding the features based on an a priori coding scheme. The literature resulted in a total of 258 selected papers. From the selected papers, a total of 835 features were selected. Based on the a priori coding scheme, the features were mapped according to the following dimensions: A) the type of features applied, B) the information source applied and, C) the type of organization that applies the features. The results show that most features focus on hard information from a transactional source from official information with a high latency. In addition, the results show that most features still relate to the traditional Altman-Z score.

All the results have been mapped on the FCRA Model, which is based on Wand and Weber [46], see Figure 3. The insights derived from this study provides a better understanding of the level on which the features are applied and where they score in the FCRA Model. This will enable further exploration and identification of features that have a low latency but still have a proper predictive power. From a practical perspective, our study provides an overview of features that can currently be applied, and which further exploration should be considered.

While we provide an integrative overview of features for FCRA, our study is not without limitations. The first limitation concerns the sampling and sample size. The sample group of features is drawn from the identified paper without considering the effectiveness of the features selected. The main reason for this choice is the fact that not all papers report on the effectiveness of the features applied. While we believe that for the purpose of this study this causes no problems, further refinement of the features selected is recommended. Additionally, our results should be further validated in practice.

References

[1] E. Mantelaers and M. Zoet, "A New Explorative Model to Assess the Financial Credit Risk Assessment," eKNOW 2018.

[2] S. Cornee and A. Szafarz, "Vive la Difference: Social Banks and Reciprocity in the Credit Market," *J. Bus. Ethics*, vol. 125, no. July, pp. 361–380, 2014.

[3] S. Balcaen and H. Ooghe, "35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems," *British Accounting Review*, vol. 38, no. 1. pp. 63–93, 2006.

[4] J. Watson and J. E. Everett, "Do Small Businesses Have High Failure Rates?," *J. Small Bus. Manag.*, vol. 34, no. 4, pp. 45–62, 1996.

[5] B. S. Ahn, S. S. Cho, and C. Y. Kim, "The integrated methodology of rough set theory and artificial neural network for business failure prediction," *Expert Syst. Appl.*, vol. 18, no. 2, pp. 65–74, 2000.

[6] T. Derriks, "A Business Process & Rules Management Maturity Model for the Dutch governmental sector," no. March, 2012.

[7] C.-F. Tsai, "Feature selection in bankruptcy prediction," *Knowledge-Based Syst.*, vol. 22, no. 2, pp. 120–127, 2009.

[8] P. Ravi Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review," *Eur. J. Oper. Res.*, vol. 180, no. 1, pp. 1–28, 2007.

[9] N. Chen, B. Ribeiro, and A. Chen, "Financial credit risk assessment: a recent review," *Artif. Intell. Rev.*, vol. 45, no. 1, pp. 1–23, 2016.

[10] A. N. Berger and L. K. Black, "Bank size, lending technologies, and small business finance," *J. Bank. Financ.*, 2011.

[11] A. N. Berger *et al.*, "Small Business Lending by Banks: Lending Technologies and the Effects of Banking Industry Consolidation and Technological Change."

[12] N. J. Kim and R. Z. Elias, "Financial statements, attestation level and lending decision by small banks," *Southwest Bus. Econ. J.*, pp. 63–72, 2008.

[13] L. Bollen, R. Meuwissen, G. M.- Faces, and undefined 2008, "Classifying major European business failures," *FST*.

[14] M. a Petersen, "Information: Hard and Soft," *Northwest. Univ. Doc. Trav.*, no. July, p. 20, 2004.

[15] M. Kano, H. Uchida, G. F. Udell, and W. Watanabe, "Information verifiability, bank organization, bank competition and bank-borrower relationships," *J. Bank. Financ.*, vol. 35, no. 4, pp. 935–954, 2011.

[16] W. W. T. Nemoto., Y. Ogura., "The Decision-Making Mechanism of Regional Financial Institutions and the Utilization of Soft Information *," *Policy Res. Institute, Minist. Financ. Japan*, vol. 9, no. January, pp. 87–116, 2013.

[17] A. N. Berger and G. F. Udell, "A More Complete Conceptual Framework for SME Finance."

[18] D. A. Carter, J. E. Mcnulty, and J. A. Verbrugge, "Do Small Banks have an Advantage in Lending? An Examination of Risk-Adjusted Yields on Business Loans at Large and Small Banks," *J. Financ. Serv. Res.*, vol. 25, pp. 2–3, 2004.

[19] H. Brown-Liburd and M. A. Vasarhelyi, "Big Data and Audit Evidence," *J. Emerg. Technol. Account.*, 2015.

[20] Y.-H. Lu, Y.-C. Lin, and Y.-L. Lin, "資料探勘技術在繼續經營疑慮意見診斷模型之應用 (Going-Concern Opinion: The Application of Data Mining Technologies)." 27-Jul-2016.

[21] E. I. Altman, G. Sabato, and N. Wilson, "The value of non-financial information in small and medium-sized enterprise risk management," *J. Credit Risk*, vol. 6, no. 2, pp. 1–30, 2010.

[22] F. Mccann and T. Mcindoe-Calder, "Firm size, credit scoring accuracy and banks' production of soft information," 2015.

[23] F. Dainelli, F. Giunta, and F. Cipollini, "Determinants of SME credit worthiness under Basel rules: the value of credit history information," *PSL Q. Rev.*, vol. 66, no. 264, pp. 21–47, 2013.

[24] J. P. Suter and I.). F. S. of B. Anderson University (Anderson, *A Quantitative Study of the Collection of Soft Information by Small Community Banks: Building Sustainable Competitive Advantage*. Anderson University

Falls School of Business, 2007.

[25] S. Angilella and S. Mazzù, "The financing of innovative SMEs: A multicriteria credit rating model," *Eur. J. Oper. Res.*, vol. 244, no. 2, pp. 540–554, 2015.

[26] E. Kirkos, "Assessing methodologies for intelligent bankruptcy prediction," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 83–123, 2012.

[27] M. Aruldoss, M. L. Travis, and V. P. Venkatesan, "A reference model for business intelligence to predict bankruptcy," *J. Enterp. Inf. Manag.*, vol. 28, no. 2, pp. 186–217, 2015.

[28] E. I. Altman, "The Prediction of Corporate Bankruptcy: A Discriminant Analysis," *J. Finance*, vol. 23, no. 1, pp. 193–194, 1968.

[29] R. O. Edmister, "An empirical test of financial ratio analysis for small business failure prediction," *J. Financ. Quant. Anal.*, vol. 7, no. 2, pp. 1477–1493, 1972.

[30] E. B. Deakin, "A Discriminant Analysis of Predictors of Business Failure," *J. Account. Res.*, vol. 10, no. 1, pp. 167–179, 1972.

[31] Y. Chen, R. J. Huang, J. Tsai, and L. Y. Tzeng, "Soft Information and Small Business Lending," *J. Financ. Serv. Res.*, vol. 47, no. 1, pp. 115–133, 2013.

[32] Y. H. Ju and S. Y. Sohn, "Updating a credit-scoring model based on new attributes without realization of actual data," *Eur. J. Oper. Res.*, vol. 234, no. 1, pp. 119–126, 2014.

[33] K. Kosmidis and A. Stavropoulos, "Corporate failure diagnosis in SMEs," *Int. J. Account. Inf. Manag.*, vol. 22, no. 1, pp. 49–67, 2014.

[34] G. F. Udell, "What's in a relationship? The case of commercial lending," *Bus. Horiz.*, vol. 51, no. 2, pp. 93–103, 2008.

[35] S. Bhattacharya and A. V. Thakor, "Contemporary banking theory," *J. Financ. Intermediation*, vol. 3, no. 1, pp. 2–50, 1993.

[36] A. C. Edmondson and S. E. Mcmanus, "Methodological fit in management field research," *Acad. Manag. Rev.*, vol. 32, no. 4, pp. 1155–1179, 2007.

[37] M. Spreen, "Rare Populations, Hidden Populations, and Link-Tracing Designs: What and Why?," *Bull. Méthodologie Sociol.*, vol. 36, no. 1, pp. 34–58, 1992.

[38] S. K. Thompson, "Adaptive sampling in behavioral surveys," *NIDA Res.Monogr*, vol. 167, no. 1046–9516 (Linking), pp. 296–319, 1997.

[39] K. Hjelm, P. Nyberg, Å. Isacsson, and J. Apelqvist, "Beliefs about health and illness essential for self-care practice: A comparison of migrant Yugoslavian and Swedish diabetic females," *J. Adv. Nurs.*, vol. 30, no. 5, pp. 1147–1159, 1999.

[40] J. H. Patrick, R. A. Pruchno, and M. S. Rose, "Recruiting Research Participants: A Comparison of the Costs and Effectiveness of Five Recruitment Strategies," *Gerontologist*, vol. 38, no. 3, pp. 295–302, 1998.

[41] J. Grunert, L. Norden, and M. Weber, "The role of non-financial factors in internal credit ratings," *J. Bank. Financ.*, vol. 29, no. 2, pp. 509–531, 2005.

[42] E. I. Altman, G. Sabato, and N. Wilson, "The value of non-financial information in small and medium-sized enterprise risk management," *J. Credit Risk*, vol. 6, no. 2, pp. 1–33, 2010.

[43] C.-C. Trönnberg and S. Hemlin, "Lending decision making in banks: A critical incident study of loan officers," *Eur. Manag. J.*, vol. 32, pp. 362–372, 2014.

[44] M. del M. Camacho-Miñano, M. J. Segovia-Vargas, and D. Pascual-Ezama, "Which Characteristics Predict the Survival of Insolvent Firms? An SME Reorganization Prediction Model," *J. Small Bus. Manag.*, vol. 53, no. 2, pp. 340–354, 2015.

[45] N. Chen, B. Ribeiro, and A. Chen, "Financial credit risk assessment: a recent review," *Artif. Intell. Rev.*, 2016.

[46] Y. Wand and R. Weber, "An Ontological Model of an Information System," *IEEE Trans. Softw. Eng.*, vol. 16, no. 11, 1990.