The Business Model DNA: Towards an Approach for Predicting Business Model Success

Markus Böhm[1], Jörg Weking[1], Frank Fortunat[1], Simon Müller[2], Isabell Welpe[2], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Garching, Germany
{markus.boehm,joerg.weking,frank.fortunat,krcmar}@in.tum.de
[2] Technical University of Munich, Chair for Strategy and Organization, Garching, Germany
{simon.mueller,welpe}@tum.de

Abstract. Business models have gained much interest in the last decade to analyze the potential of new business ventures or possible innovation paths of existing businesses. However, the business model concept has only rarely been used as basis for quantitative empirical studies.
This paper suggests the concept of a Business Model DNA to describe the characteristics of specific business models. This concept allows to analyze business models in order to identify clusters of business models that outperform others and calculate future prospects of specific business models.
We used 181 startups from the USA and Germany and applied data mining techniques, i.e. cluster analysis and Support Vector Machines, to classify different business models in regards to their performance.
Our findings show that 12 distinct business model clusters with different growth expectations and chances of survival exist. We can predict the survival of a venture with an accuracy of 83.6 %.

Keywords: Business Model; Success Prediction; Data Mining; Cluster Analysis; Support Vector Machine

1 Introduction

Business models (BMs) have established in different research communities, like management science and information systems, whereas in practice they are seen as vital for business success [1, 2]. Furthermore, a shift from traditional BMs to electronic ones took place in the last three decades [3, 4]. There are many definitions and frameworks available on the concept today [1].

Business modelling gained significant importance in startup communities, too [2]. Startup firms are a driver of economic growth, innovation and employment opportunities [5]. Unfortunately the failure rate of startups is very high, with estimates ranging from 50% to more than 83% [6, 7]. Why some new ventures fail, while others succeed, is one of the central questions not only for entrepreneurship research, but also for possible entrepreneurs [2, 8]. Scholars agree that current research is still lacking methods to predict firm success [2, 9-11]. Additionally, most of the previous studies

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Böhm, M.; Weking, J.; Fortunat, F.; Müller, S.; Welpe, I.; Krcmar, H. (2017): The Business Model DNA: Towards an Approach for Predicting Business Model Success, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1006-1020

are based on qualitative research [2, 9]. This study proposes a new method for predicting firm success taking into account quantitative measures.

Through the emergence of more powerful data analysis tools and the growing amount of available data, it is possible to use data mining to find meaningful patterns in datasets [12]. Data mining is concerned with making sense of large amounts of data and finding patterns that are difficult to find manually [13].

The goal of this paper is to combine data mining approaches (i.e. cluster analysis and Support Vector Machines (SVM)) with the BM concept to predict the chances of success for startups. This allows better informed, empirically backed strategic and investment decisions [14]. To operationalize this approach, we suggest the BM DNA – in analogy with the human genome – as a concept to describe the characteristics of a specific BM based on various factors. The proposed instance of the BM DNA has been drawn from the 55 BM patterns based on Gassmann, Frankenberger and Csik [15].

The remaining paper is structured as follows. The second chapter describes related work including relevant BM literature. The third chapter elaborates the method that is applied in order to evaluate BMs of startup firms. The fourth chapter illustrates the dataset that has been used. Results of the cluster analysis and the different models of a SVM for classifying BMs are shown in the fifths chapter. The sixth chapter discusses these results and limitations of the study. The final chapter concludes with a short summary, contributions and aspects for future research.

2 Related Work

More than 60 years ago, Drucker [16] defined the term BM as the answer to “who is the customer, what does he value and how can you make money from it”. With the rise of the internet and digital firms the concept has gained more attention in research and practice [1, 10, 17]. This has led to various definitions of BMs in different research streams [18, 19].

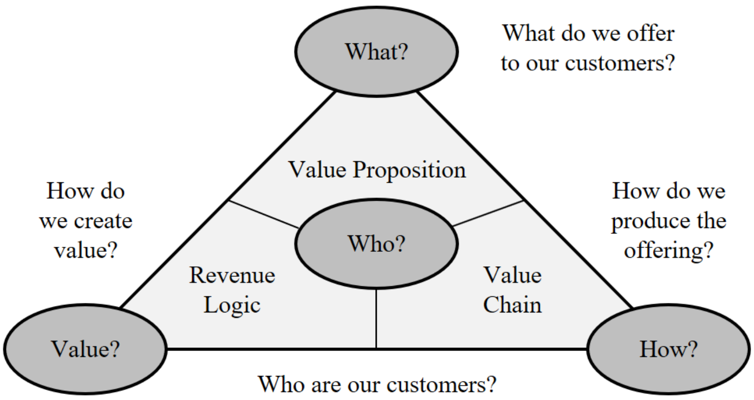

Figure 1. The magic triangle of business models [15]

Drawing from common BM definitions, Gassmann, Frankenberger and Csik [15] have developed a BM framework that consists of four central dimensions: the *Who*, the *Wha*t, the *How*, and the *Value* (see Figure 1). This framework builds the foundation of their 55 BM patterns, which they have identified by analyzing 250 BMs along the four dimensions of the framework. These 55 patterns describe the core of how a BM works. Gassmann, Frankenberger and Csik [15] assume that 90% of the BMs that exist today can be broken down into the same 55 patterns.

Classification of BMs has a long history in entrepreneurship and e-commerce research [20]. Previous studies have used classifications of BMs in empirical research to analyze the influence of BMs on enterprise performance [9]. Lambert and Davidson [9] reviewed research from 1996 to 2010 and found 40 papers that investigate this relationship. Studies that focused on e-commerce conclude that firms should concentrate on interaction platforms for facilitating online transactions and advertising as major revenue stream to be successful. Other studies show that a strategical focus on BM innovation leads to a higher profitability. Additionally, the usage of novelty-centered BMs is related to enterprise success. Moreover, some studies demonstrate that if a BM can easily be transferred to other markets, the firm will more likely be successful [9].

Weill, Malone and Apel [21] distinguish BMs along two dimensions, asset types (financial, physical, intangible, human) and asset rights (creator, distributor, landlord, broker). This leads to a matrix of 16 theoretically possible BM clusters, of which 14 are legal. This framework was used to classify more than 10'000 firms, publically traded on U.S. stock exchanges in order to perform a stock market analysis. They discovered that innovative manufactures are most valued by the stock market. These manufactures are defined as organizations that invest more in research and development than the industry average [21].

Spiegel, Abbassi, Zylka, Schlagwein, Fischbach and Schoder [2] focused on the success of early stage startups using empirical qualitative research. They found that the founders' social capital is crucial to receive funding in this situation. Founders with a better professional social network are able to develop a better BM using their contacts for information and status benefits [2].

However, only few of these studies have focused on quantitative empirical research to quantitatively analyze whether some BMs perform better than others [2, 9]. Scholars agree that future work should aim for the better understanding of relationship of BMs and firm performance [2, 9-11].

3 Research Method

To analyze BM patterns of successful startups, we rely on the Mattermark dataset which is a collection of information about startup firms. In order to avoid a bias towards US or German firms, we drew a random sample of 75 US and 75 German startup firms. For this sample we collected additional information to describe the BM by manually searching the web. Additional information sources included firm websites, press reports and interviews with founders that were publically available.

As the Mattermark dataset is strongly biased towards successful startups we identified additional failed ventures from Crunchbase-Insights, Deadpool and autopsy.io. We sent out a survey to a total of 210 firms and 309 founders asking for additional information to match the data available in the Mattermark dataset. In this way, we were able to include 31 failed startups with sufficient information on financing, revenues, competition and innovativeness of the product.

In the first step of the analysis the BM DNA is built for each of the 181 firms. Similar to the DNA as a molecule that carries most of the genetic information of living organisms, the BM DNA should be a representation of a BMs characteristics. Hence, it should describe the essence of a business in a precise way that allows comparison to other BMs.

In order to build a BM DNA, we draw on the 55 BM patterns developed by Gassmann, Frankenberger and Csik [15]. As 90% of the BMs that exist today can be broken down into the same 55 patterns [15], these BM patterns serve as a sufficient basis for describing the BM DNA. For each of the 55 patterns we used a binary variable to indicate whether a firm uses this pattern or not. In this sense, the BM DNA is a vector that indicates the patterns a certain BM applies. Figure 2 is an exemplary visual representation of the BM DNA. We manually evaluated each firm with regards to the BM patterns it applies. This evaluation and coding process was solely done by one person to ensure a consistent coding of the BM DNAs.

1	2	3	4	5	6	...	49	50	51	52	53	54	55
0	1	1	0	0	1		1	0	1	1	0	0	1

Figure 2. Representation of an exemplary BM DNA

The second step of the data analysis is a cluster analysis based on the BM DNAs. This is not only done to result in meaningful clusters of BMs with different growth perspectives, but also to improve the results of the SVM [22]. We applied the k-means clustering algorithm according to Jain, Murty and Flynn [23] with squared Euclidean distances. The algorithm was used in an iterative process with different numbers of clusters (k) as input factor. The final number of clusters has then been determined with regards to two criteria, the maximum distance between the clusters (I) and the meaningfulness of the clusters (II).

In the third step we developed a metric to evaluate the success of a venture. Therefore, we used both the survival of a firm and its revenue growth. In order to have a comparative metric for revenue growth, we measure the growth relative to the actual revenue (see Figure 4 in the next chapter). The less revenue a firm generates the more revenue growth it needs to generate in order to be evaluated successful.

As a fourth step a SVM is used to classify BMs according to their growth perspectives and whether they are successful or unsuccessful. SVMs use a nonlinear mapping to transform input data (training data) into high dimensional data. The method searches for an optimal, maximum marginal separating hyperplane [24]. This hyperplane is based on support vectors which can be seen as a small subset of the training data [22]. The SVM has been chosen since both neural networks and SVMs have shown to deliver satisfying results in similar studies [25]. An initial comparison of these two techniques using our data indicated better results for SVMs. Thus, SMVs are used in this study with the following information as input:

- BM Information: BM DNA, cluster, scope, focus (B2C or B2B), industry, physical assets, firm age
- Involved people: industry/ foundation experts, investors, founding team size, education of founders, location (Country & City)
- Startup idea: closeness to science and patents, competition and innovativeness

The dataset is randomly split into 80% of the data for training and 20% for testing since this often leads to optimal results [26]. The training data is used to produce the model. This model is then fed with the test data.

The performance of the model is measured by an accuracy, kappa and area under the curve (AUC). The kappa value shows the difference between the calculated solution and a random solution. A kappa value of 0 would indicate a random classification whereas a kappa value of 1 stands for a perfect solution [27]. The AUC measure indicates as well how the model performs in comparison to a random model. The random model would have an AUC of 0.5, whereas an AUC of 1 would indicate a perfect classifier [28].

4 Dataset

The dataset includes 181 firms in total, 31 failed firms resulting from the survey and 150 active firms from the Mattermark dataset. Eighteen of the 31 failed firms were founded in the USA. The databases have been accessed in Mai 2015. The majority of firms is founded between 1999 and 2015. The initial coverage of BM patterns resulting from the manual coding is shown in Figure 3. Almost all firms apply the BM pattern *Digitalization* (11) which stands for digital products or services [15]. However, the dataset is not focused on digital industries.

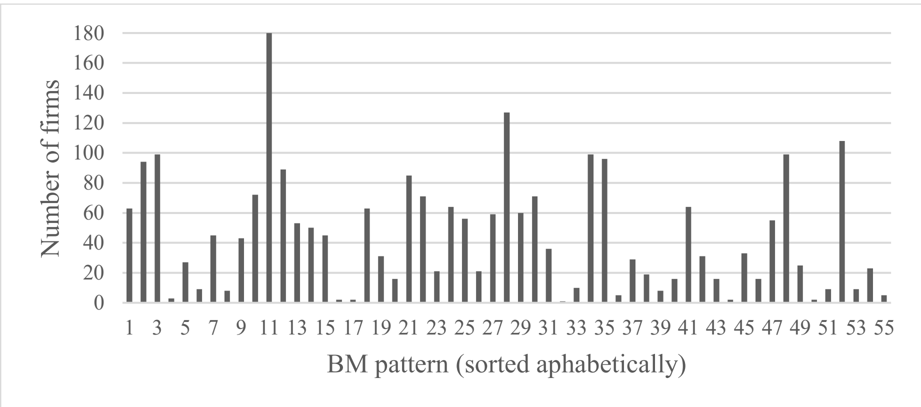

Figure 3. Initial coverage of BM pattern

To get to a meaningful result for the success measure, the data for the USA and Germany had to be treated differently due to the fact that the German firms did not satisfy the same growth criteria than the US firms. This can be linked to the smaller number of average investors and funding the German firms had. Average funding of US startups (238.2 $ M) is more than 3 times higher than the funding German startups received (58.2 $ M). Additionally, German startups had three investors on average while startups from the USA had almost 9 on average.

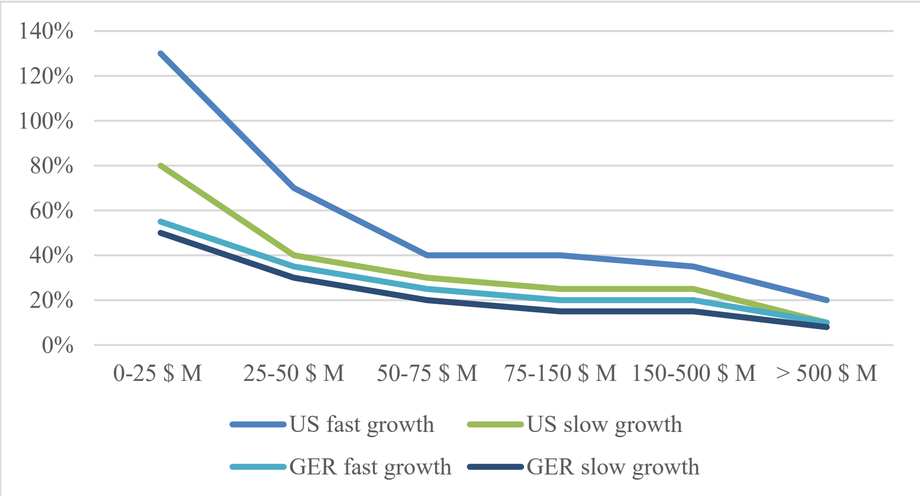

Figure 4. Revenue growth thresholds adapted from [29]

Less funding means fewer possibilities to invest in rapid growth and, therefore, the two regions are not comparable with the same growth percentages. The thresholds were determined in accordance to Maltz and Saljoughian [29] and adjusted for German firms. Figure 4 shows the thresholds that separate the different firms into fast growing and slow growing firms.

5 Predicting Business Model Success

5.1 Business Model Clusters

In the following the characteristics of each of the 12 identified clusters are elaborated. One cluster resembles firms that have a similar combination of BM patterns (BM DNA). Table 1 shows which BM patters are mainly applied by the firms in the respective cluster.

Table 1. Mainly applied business model patterns by cluster

Cluster	*Mainly applied BM patterns*
Freemium Platforms	Freemium and Platforms (Orchestrator, Two-Sided Market or Long Tail)
Experience Crowd Users	Experience Selling, Crowdsourcing and Leverage Customer Data
Long Tail Subscribers	Long Tail and Subscription
Affiliate Markets	Aikido, Affiliation and Platforms (Two-Sided Market, Orchestrator or Long Tail)
Mass Customizing Orchestrators	Mass Customization, Layer Player, Orchestrator and Two-Sided Market
Innovative Platforms	Aikido, Two-Sided Market, Orchestrator and Revenue Sharing
E-Commercer	E-Commerce and Direct Selling
E-Commerce Affiliates	E-Commerce, Affiliate and Long Tail
Add-On Layers	Add-On, Layer Player and Subscription
Crowdsourcing Platforms	Aikido, Crowdsourcing, Customer Loyalty and Platforms (Two-Sided Market or Orchestrator)
Customized Layers	Subscription and Mass Customization
Hidden Revenue Markets	Hidden Revenue, Two-Sided Market, Affiliation and Long Tail

Table 2 indicates the clusters, the number of startup firms in a cluster (n) and their chances of success in terms of survival, fast growth and slow growth. The numbers are to be interpreted as the percentage of firms in the cluster that met the success criteria described in section 3. Thus, firms in the category slow growth also include those firms that managed to grow fast and survival category also includes those firms that managed to grow slowly or fast. The number of members per cluster is well balanced with the fewest members in cluster 0 and 5 with 6. Cluster 7 has the most members with 29.

The *Freemium Platform* cluster mostly includes firms that are using a *Freemium* model or a platform (e.g. *Orchestrator*, *Two-Sided Market* or *Long Tail*). The former offers a free basic service to attract a broad customer base. Revenues are generated by additional chargeable offers. *Orchestrator* coordinate value activities of various firms to offer their customers an aggregated product. The BM of the *Two-Sided Market* serves different customer groups with one platform. The platform is only interesting for any of the two customer groups if the other one is present in a sufficient number as well. For this reason, free services are usually offered to one group of customers. The basis of both BM patterns is a mediation platform. The *Long Tail* pattern is used when the profits for many small payments are achieved by a small margin. Concluding, this cluster includes firms that aggregate the services of different providers and offer a free basic service on their platform as well as a supplementary service. The number of firms in this cluster is rather low (n=6). While all the firms in the sample survive, only half of the firms exhibit strong sales growth. Thus, it is assumed that only a few customers are willing to pay for additional services.

Table 2. Successful business models by cluster

Cluster	*n*	*Fast Growth*	*Slow Growth*	*Survival*
Freemium Platforms	6	50%	50%	100%
Experience Crowd Users	11	18%	27%	64%
Long Tail Subscribers	16	75%	75%	94%
Affiliate Markets	23	52%	65%	78%
Mass Customizing Orchestrators	12	50%	67%	83%
Innovative Platforms	6	67%	67%	100%
E-Commercer	14	43%	57%	57%
E-Commerce Affiliates	29	45%	45%	76%
Add-On Layers	20	45%	60%	90%
Crowdsourcing Platforms	11	73%	91%	91%
Customized Layers	13	54%	62%	69%
Hidden Revenue Markets	20	25%	25%	75%

The *Experience Crowd Users* cluster includes firms that pursue *Experience Selling* and *Crowdsourcing*. In addition, firms in this cluster *Leverage Customer Data*. As part of the *Experience Selling* BM, experience of customers is the key part of the customer value proposition. The goal of the offer is a unique experience. As part of the *Crowdsourcing*, central activities of the value creation are transferred to the crowd, either to the general public or a selected group. This makes it possible to use the potential of large user groups for the firm. The *Leverage Customer Data* pattern is characterized by drawing advantage from customer data. This can be achieved by offering tailored advertising, for example. However, this cluster has the second worst survival rate (64%) and the worst success rate (18%). Our findings suggest that it is very difficult to survive with this BM. Even if startups survive, the growth opportunities are poor. Thus, this cluster is more suitable for niche markets.

The *Long Tail Subscriber* cluster mainly relies on the *Long Tail* and the *Subscription* pattern as an integral part of its strategy. The latter refers to a contractually fixed periodic payment of the user to the provider. In return, the customer can use the offering during a certain period of time. This combination in digital goods is very successful, as duplication of information or provision of software solutions cause almost no additional costs. Therefore, products can be offered very cheap and greater profits can be achieved by many smaller payments over a longer period. Since 94% in this cluster survived and 75% of firms experienced a fast growth, it can be seen as a very successful one.

In addition to the affiliate model, entities within the *Affiliate Markets* cluster utilize the classic patterns of platforms (i.e. *Two-Sided Market*, *Orchestrator* and *Long Tail*). In addition, the *Aikido* pattern is employed. Under the *Affiliation* BM, client referrals to third parties are rewarded with commissions. Price comparison portals are an example of the combination of the affiliate model with a platform. *Aikido* comprises BMs that concentrate on something different than competitors in the same industry. An example is a transfer of BMs that have proven to be successful in other industries. The survival rate of this cluster is average.

The *Mass Customizing Orchestrators* connect the *Mass Customization* pattern with *Layer Player*, the *Orchestrator* and the *Two-Sided Market* pattern. *Mass Customization* means that products, although produced in mass production can be individualized to a certain degree though a variety of options. If a solution for a particular part of the value chain is offered in various industries the *Layer Player* pattern is used. This cluster combines an individual offer with a platform on which the partial offers are available for individual partners. It has a survival rate of 83% and a fast growth rate of 50%. Thus this cluster is quite convenient to grow rapidly.

The next cluster, *Innovative Platforms* focuses besides the classic patterns of a platform such as *Two-Sided Market*, *Orchestrator* and *Revenue Sharing*, mainly on the *Aikido* pattern. Thus, firms of this cluster are trying to integrate new ideas in the BM of platforms. This can be seen as quite successful because all of the firms of the sample have survived until now and two-thirds are also growing fast.

The *E-Commercer* cluster is characterized by the pattern of *E-Commerce* and *Direct Selling*. Entities in *E-Eommerce* offer products or services through the internet. *Direct Selling* means that no intermediaries are used to sell products. This cluster has the lowest survival rate and the second lowest rate of success. This is because the e-commerce market is already saturated. Entering the market is very difficult for young firms without innovative BMs due to strong price competition. An example firm of this cluster is Zalando.

The cluster *E-Commerce Affiliates* performs slightly better than the *E-Commercer* cluster. Besides *E-Commerce* patterns, it follows the *Affiliation* and *Long Tail* pattern. About 3 out of 4 firms survive in this cluster. Only 45% can generate strong growth. By aggregating the offers of many suppliers, a wide selection is created for the customer. Money is generated through links to websites of actual shops.

The *Add-On Layer* uses the *Add-On* pattern, *Layer Player* and *Subscription*. Here, a basic offer is provided relatively cheap. A surcharge must be paid to use more options. This pattern is often used in software as a service (SaaS) products. 90% of firms that use this BM have survived so far. However, only 45% grow fast. This can be attributed to the fact that for many services, the majority of the users just employ basic versions.

Aikido, *Crowdsourcing*, *Customer Loyalty* and the aforementioned platform patterns are used by the cluster *Crowdsourcing Platforms*. *Customer Loyalty* tries to lock-in the customer through incentives such as a bonus scheme or rewards for repeated use. Especially for crowdsourcing, it is extremely important to develop an active community. Therefore, the combination of these two patterns is quite useful. Likewise, a platform for exchange is required for crowdsourcing. The success ensures that users of this BM are right. Nine out of ten firms survive and nearly three out of four grow fast. One potential reason could be that a strong transformation from a pure consumer to dialogue with firms has taken place in recent years. For this reason, many customers are willing to invest time and energy in crowdsourcing campaigns.

The *Customized Layers* also use the *Subscription* and the *Mass Customization* model. The combination of *Layer Player* and *Mass Customization* is very common, especially in the SaaS industry. In this cluster, a hull software is used in many industries but tailored to every firm. We find that slightly less than 75% of the firms survive in this cluster. Additionally, only 54% exhibit strong growth.

The *Hidden Revenue Markets* rely on the *Hidden Revenue* and the *Two-Sided Market* pattern. In addition, *Affiliation* and *Long Tail* are included. In the *Hidden Revenue* pattern there is usually one offer free of charge for one side of a two-sided market. The other group of customers have to pay. The combination of *Long Tail* and *Hidden Revenue* suggests a high volume of transactions. Otherwise it would not be possible to sustain the offer with low payments. A famous example firm of this cluster is the Telegram Messenger. The cluster with this BM does not perform very well. Although three out of four firms survive, only one out of four grows fast.

The 12 clusters show some significant differences. Even though some clusters like the *Freemium Platforms* had a high survival rate, the rate of success was considerably lower compared to other clusters. The same can be seen in regards to the *Innovative Platforms* cluster as well as the *Add-On Layers* cluster. The most consistent cluster was the *Crowdsourcing Platform* cluster which had a survival rate of 91% and a fast growth rate of 73%. This means firms of this cluster that survive tend to grow fast and strong. The *Experience Crowdusers* cluster has a low survival rate of 64% as well as a low success rate of 18%. This cluster can be seen as a niche market. The *E-Commercer* cluster also shows that this is a market with strong competition as most firms do not make it. The few ones that survive grow fast and big. This observation can be seen in the e-commerce market which is dominated by a few big players. The *Long Tail Subscribers* use an interesting BM that combine a big customer base paying a relative low price with a subscription model. In this way the firms are able to attract enough customers to survive and to grow fast. This is as well shown by their survival and success rate of 94% and 75%. Overall it can be seen, that of the 12 clusters 7 are above the 50% survival threshold and only the *Experience Crowdusers* and the *Hidden Revenue Markets* clusters are well below this threshold.

5.2 Success Classification Models

Table 3. Accuracy of success classification

	Model	*calculated*	*true*		*class*	*Kappa*	*Accuracy*	*AUC*
			yes	*no*	*precision*			
1	Fast growth	yes	21	6	77.78%	0.312	66.59%[1]	0.744
		no	12	15	53.57%			
2	Slow growth	yes	26	9	74.29%	0.339	66.8%[1]	0.734
		no	8	12	60.00%			
3	Survival	yes	46	2	95.83%	0.673	83.6%[1]	0.899
		no	2	5	71.43%			
4	Fast growth only BM	yes	17	7	70.83%	0.153	58.4%[1]	0.629
		no	17	14	45.16%			

The SVM is used with different models to classify ventures according to their growth perspectives and survival. Table 3 shows the results of the different classification models. The first model separates firms between fast growing and not fast growing. For example, the model claims that 27 firms are fast growing. However, only 21 of these

[1] Based on 50-50 weighted sample

27 firms are actually classified as fast growing. In other words, their growth relative to their revenue lies above the threshold depicted in Figure 3. The model is able to correctly classify two thirds of the given test data.

The second model takes into account whether a firm is slowly growing or not slowly growing. Hereby, it shows a similar accuracy like the first model. The third model, however, is able to correctly classify 83.6% of given startups in regards to their survival or not. The fourth model is as well concerned with the separation of fast growing firms and not fast growing ones. However, the fourth model uses only information about the BM DNA as an input.

The accuracy considerably varies in the different models. The survival model is the most accurate one as it has a high accuracy as well as a high Kappa measure of 0.673 which is classified as substantial. The AUC measure is also promising with a value of 0.899. The other models lack in accuracy. Furthermore, the success models, models one and two, have Kappa values of around 0.3, which is classified as fair [27]. The AUC values were respectable for both success models though as they were 0.744 and 0.734 respectively.

The fourth model proved that BM information by itself has only very weak classification ability. This means that it is not possible to achieve a high classification accuracy without a combined approach. The low kappa value of 0.153 and AUC value of around 0.6 underlines this statement.

6 Discussion

The paper discusses two data mining approaches for classifying startup BMs in terms of their success by building on a new concept called BM DNA. The first result, arising from a cluster analysis, shows currently promising or not promising BM clusters. As part of the second result a SVM is introduced for classifying BMs as successful or unsuccessful.

The k-means clustering algorithm is used in this study since it has shown to be very efficient [23]. However, the algorithm shows some downsides. The number of clusters is an input factor, the outcome is depended on the initial solution, it is sensitive to outliers and can end up in local optimal solutions instead of global ones [23, 30]. In addition, the Euclidian distance used with binary data can be seen as problematic [23]. The iterative usage mitigates these disadvantages. Alternatively, a hieratical agglomerative clustering algorithm can be used and is seen as future research.

The clusters achieved a satisfactory result even though some similar patterns correlate with different clusters. These patterns are applied in many companies. On the other hand, some patterns, like *Cash Machine*, *Trash-to-Cash*, *Target the Poor*, *White Label* or *Ultimate Luxury*, do not show a strong correlation with any pattern. There were not enough firms that use these patterns in the sample.

There are different models in the literature aiming to classify BMs in regards to survival and not survival (Table 3). The survival model (3) correctly predicts 95.83% of the successful firms and 71.43% of the unsuccessful firms without considering economic anomalies. This accuracy is as good as or even better than most studies of

Business Failure Prediction (BFP) for new ventures. Lussier [31] was able to match 73% of the failed and 65% of the successful firms by using logistic regression.

This study used SVM as classification method. Different methods are used in literature. Gartner, Starr and Bhat [32] used multiple discriminant analysis and were able to correctly classify 85% of the firms which is similar to this approach. Marom and Lussier [33] achieved an accuracy of 85.4% which is also comparable. However, our model is superior in classifying successful businesses as Marom and Lussier [33] were only able to classify the successful ones with an accuracy of 84%. Ciampi and Gordini [34] were able to correctly classify 68.4% of their firms. Ecer achieved results of 91.18% and 88.24% for neural networks and SVMs respectively. Hence, both techniques have a high prediction ability [25]. Ecer [25] achieved a better accuracy using neural networks while Olson, Delen and Meng [24] had the best result with decision trees. Wilson and Sharda [35] used the same financial ratios as Altman since they wanted to compare neural networks and multiple discriminant analysis. Here, multiple discriminant analysis was outperformed by neural networks in every test. However, in this study, the prediction ability of the SVM proved to be superior for the survival model in comparison to other methods in literature.

There are some limitations of SVMs. Information about the founders and their prior knowledge of the industry and founding, the degree of innovativeness of the new firm as well as competition, founding team size, and patents improved the model accuracy [7, 31, 36-38]. While the variables did produce a satisfactory model, the problem of transparency and transportability, the inherent issue of all SVMs, still exists [24]. With SVMs, it is not possible to show how much each variable is influencing the outcome in a numerical way like with a logistic regression. Unfortunately, it is also not possible to visualize the model since it has too many dimensions for producing a graphical representation. Other authors used techniques like ANOVA, principal component analysis (PCA) or correlation matrices to preselect the variables they used in their models [39, 40]. However, these approaches neglect the interaction of the independent variables among themselves and can lead to a worse selection of variables. Thus, SVM has been used in this study and delivers very good results compared to other studies.

While the survival model was able to deliver a good prediction accuracy, models used on mature firms that include profound financial ratios are superior in accuracy. The model by Altman was able to correctly classify 95% of the cases. Edminster also used a number of financial ratios in his multiple discriminant analysis model and had a 92% classification success rate on small firms [41].

However, the focus of this paper is on startup firms. The rate of failure in startups is considerably higher than in mature firms [2]. Additionally, in early stage startups, it might be difficult to get profound financial numbers. Thus, a success model with an accuracy of almost 85% is a very good classification model for startups.

The proposed model in this paper is a considerable contribution for startup classification and success prediction. However, this study is a proof of concept only. Our study showed that the used method is able to correctly classify startups concerning their overall success. More data is needed to clearly verify, enhance and refine the model.

7 Conclusion

This paper suggests a BM DNA as means to describe the specific characteristics of certain BMs. It demonstrated the applicability of this concept by drawing from the 55 BM patterns based on Gassmann, Frankenberger and Csik [15] to describe the characteristics of a sample of 181 firms.

The BM DNA was then used in a cluster analysis. 12 different BM cluster were created including their chances for survival, for slow growth and for fast growth. These 12 cluster can be used in practice to make an assumption for growth perspectives of an actual startup. Some BM clusters seem to have better future prospects than others.

In a next step a SVM has been used to estimate a BMs survival and growth perspectives. The classification of survival worked fairly well. It is a major contribution for research since comparable studies do not achieve similar accuracies. Furthermore, an advanced model should be used in practice by entrepreneurs and investors as a step of business model evaluation.

However, the classification of slow and fast growth as a means to evaluate the success of a BM was not very reliable. Future research should investigate in additional classification algorithms (1), apply the model on a larger dataset (2) and extend the BM DNA with additional information (3). Our fourth model shows that this is needed to correctly classify BMs concerning their growth perspectives.

While future versions of the BM DNA serve as a good predictor of BM success, it is still a reflection of data from the past and, thus, past success factors. Innovation and consequently the chances for exceptional success will remain a creative process of identifying new approaches and promising re-combinations. However, predictions based on the BM DNA may serve as a sounding board for entrepreneurs and investors to critically reflect on specific BMs and make purposeful decisions.

8 Acknowledgements

This paper originated in the context of the study “Business Models in the Digital Economy – A Study of the German Innovation System” from the Commission of Experts for Research and Innovation [42]. Parts of the paper at hand have been used in this study.

References

1. Wirtz, B.W., Pistoia, A., Ullrich, S., Göttel, V.: Business Models: Origin, Development and Future Research Perspectives. Long Range Planning 49, 36-54 (2016)
2. Spiegel, O., Abbassi, P., Zylka, M.P., Schlagwein, D., Fischbach, K., Schoder, D.: Business Model Development, Founders' Social Capital and the Success of Early Stage Internet Start-Ups: a Mixed-Method Study. Information Systems Journal 26, 421-449 (2015)
3. Baden-Fuller, C., Morgan, M.S.: Business Models as Models. Long Range Planning 43, 156-171 (2010)

4. Magretta, J.: Why Business Models Matter. Harvard Business Business Review 80, 86-92 (2002)
5. Yankov, B.: Overview of Success Prediction Models for New Ventures. International Conference on Automatics and Informatics 12, 13-16 (2012)
6. Laitinen, E.K.: Prediction of Failure of a Newly Founded Firm. Journal of Business Venturing 7, 323-340 (1992)
7. Wetter, E., Wennberg, K.: Improving Business Failure Prediction for New Firms: Benchmarking Financial Models with Human and Social Capital. Journal of Private Equity 12, 30-37 (2009)
8. Cooper, A.C.: Challenges in Predicting New Firm Performance. Journal of Business Venturing 8, 241-253 (1993)
9. Lambert, S.C., Davidson, R.A.: Applications of the Business Model in Studies of Enterprise Success, Innovation and Classification: An Analysis of Empirical Research from 1996 to 2010. European Management Journal 31, 668-681 (2013)
10. Al-Debei, M.M., Avison, D.: Developing a Unified Framework of the Business Model Concept. European Journal of Information Systems 19, 359-376 (2010)
11. George, G., Bock, A.J.: The Business Model in Practice and its Implications for Entrepreneurship Research. Entrepreneurship Theory and Practice 35, 83-111 (2011)
12. Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P.: Analytics: The Real World Use of Big Data. IBM Institute for Business Value (2012)
13. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. AI Magazine 17, 37 (1996)
14. McAfee, A., Brynjolfsson, E.: Big Data: The Management Revolution. (cover story). Harvard Business Review 90, 60-68 (2012)
15. Gassmann, O., Frankenberger, K., Csik, M.: The Business Model Navigator: 55 Models That Will Revolutionise Your Business. Pearson, Harlow (2014)
16. Drucker, P.: The Practice of Management. Harper & Row, New York (1954)
17. Chesbrough, H., Rosenbloom, R.S.: The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-Off Companies. Industrial and Corporate Change 11, 529-555 (2002)
18. Krcmar, H., Böhm, M., Friesike, S., Schildhauer, T.: Innovation, Society and Business : Internet-based Business Models and their Implications. 1st Berlin Symposium on Internet and Society, 1-33 (2011)
19. Zott, C., Amit, R., Massa, L.: The Business Model: Recent Developments and Future Research. J Manage 37, 1019-1042 (2011)
20. Weiner, N., Renner, T., Kett, H.: Geschäftsmodelle im Internet der Dienste: Aktueller Stand in Forschung und Praxis. Fraunhofer Verlag, Stuttgart (2010)
21. Weill, P., Malone, T.W., Apel, T.G.: The Business Models Investors Prefer. Mit Sloan Manage Rev 52, 16-20 (2011)
22. Çomak, E., Arslan, A.: A new Training Method for Support Vector Machines: Clustering k-NN Support Vector Machines. Expert Systems with Applications 35, 564-568 (2008)
23. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. ACM Computing Surveys 31, 264-323 (1999)
24. Olson, D.L., Delen, D., Meng, Y.: Comparative Analysis of Data Mining Methods for Bankruptcy Prediction. Decision Support Systems 52, 464-473 (2012)

25. Ecer, F.: Comparing the Bank Failure Prediction Performance of Neural Networks and Support Vector Machines: The Turkish Case. Ekonomska istraživanja 26, 81-98 (2013)
26. Wang, J., Neskovic, P., Cooper, L.N.: Training Data Selection for Support Vector Machines. International Conference on Natural Computation, 554-564. Springer (2005)
27. Viera, A.J., Garrett, J.M.: Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine 37, 360-363 (2005)
28. Pruessner, J.C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D.H.: Two Formulas For Computation of the Area Under the Curve Represent Measures of Total Hormone Concentration Versus Time-Dependent Change. Psychoneuroendocrinology 28, 916-931 (2003)
29. Techcrunch, How Fast Should You Be Growing?, http://techcrunch.com/2013/08/24/how-fast-should-you-be-growing (Accessed: 21.07.2016)
30. Rai, P., Singh, S.: A Survey of Clustering Techniques. International Journal of Computer Applications 7, 1-5 (2010)
31. Lussier, R.N.: A Nonfinancial Business Success versus Failure Prediction Model for Young Firms. Journal of Small Business Management 33, 8-20 (1995)
32. Gartner, W.B., Starr, J.A., Bhat, S.: Predicting New Venture Survival: An Analysis of "Anatomy of a Start-Up." Cases from Inc. Magazine. Journal of Business Venturing 14, 215-232 (1999)
33. Marom, S., Lussier, R.N.: A Business Success Versus Failure Prediction Model for Small Businesses in Israel. Business and Economic Research 4, 63-81 (2014)
34. Ciampi, F., Gordini, N.: Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. Journal of Small Business Management 51, 23-45 (2013)
35. Wilson, R.L., Sharda, R.: Bankruptcy Prediction Using Neural Networks. Decision Support Systems 11, 545-557 (1994)
36. Hyytinen, A., Pajarinen, M., Rouvinen, P.: Does Innovativeness Reduce Startup Survival Rates? Journal of Business Venturing 30, 564-581 (2015)
37. Song, M., Podoynitsyna, K., van der Bij, H., Halman, J.I.M.: Success Factors in New Ventures: A Meta-analysis. Journal of Product Innovation Management 25, 7-27 (2008)
38. Cassar, G.: Industry and Startup Experience on Entrepreneur Forecast Performance in new Firms. Journal of Business Venturing 29, 137-151 (2014)
39. Lee, K.C., Han, I., Kwon, Y.: Hybrid Neural Network Models for Bankruptcy Predictions. Decision Support Systems 18, 63-72 (1996)
40. Dimitras, A.I., Slowinski, R., Susmaga, R., Zopounidis, C.: Business Failure Prediction Using Rough Sets. European Journal of Operational Research 114, 263-280 (1999)
41. Edmister, R.O.: An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. Journal of Financial and Quantitative Analysis 7, 1477-1493 (1972)
42. Müller, S., Böhm, M., Schröer, M., Bakhirev, A., Baiasu, B., Krcmar, H., Welpe, I.: Geschäftsmodelle in der Digitalen Wirtschaft, Studien zum Deutschen Innovationssystem. Berlin: EFI (2016)