Construction of a fuzzy model for the success prediction of hi-tech companies with a short history

Marie Jedličkova[a*], Petr Kutnar[a]

[a] *Brno University of Technology, Faculty of Business and Management, Kolejní 2906/4, Brno 612 00, Czech Republic*

Abstract

Purpose of the article There is a significant number of new high technology startup companies being built and failing on current markets every day. High-tech startups, as often abbreviated, are new companies presenting cutting edge technology. They possess the potential to change the way humankind lives and thinks. One of the essential aspects of a new startup is how to fund it. The main source of financing startups are individual investors by venture capital, another possibility is IPO, also known as public offering. Due to the nature of high-tech business it is nearly impossible to predict the success or failure of the examined startup. Investors (or venture capitalists) are often presented with many startups to choose from to fund and therefore how to value a new startups is one of the important questions still left for research to answer. This paper deals with the possibility of using fuzzy logic to bring a tool to evaluate investigating options using key factors of said startups by predicting its success or failure.

Methodology/methods The focus of this paper in the area of startups is mainly on the cleantech firms, which are companies using renewable resources and energies to provide an added value. The main goal of this paper is to create a viable and consistent fuzzy model that shows promising results for the prediction of new notions on the studied problematic. The method used to meet said goals was the usage of fuzzy software tool and performing several tests on studied data set.

Scientific aim Studied topic aims to interconnect the research field of economics with mathematical methods and computational approaches to produce a scalable and reusable software tool for business success prediction.

Findings Each iteration showed results that were used as an input to improve the next model to show more promising results. This paper concludes with a fuzzy model that is consistent enough for studied data set and can be further used for the prediction of clean tech business's success or failure.

Conclusions The fuzzy approach is an appropriate method for studied data and performed tests of designed fuzzy model show consistent results in prediction of business success in cleantech startup firm with a very short history of existence and limited data to prove their overall health.

Keywords: fuzzy logic, cleantech, startup, business success prediction, fuzzy model consistency

JEL Classification: M21, C53, C63

* Corresponding author.
E-mail address: xpjedli02@vutbr.cz

Introduction

The importance of the proper funding of high-tech startups with high potential is undeniable. With the proper funding and guidance from the investor a microscopic firm can evolve into a market changing business. One of the many sectors of high-tech startups is so called clean technology firms. These are companies with a focus on environmentally friendly technologies that will help to ease global warming and help toward a more sustainable way that humankind uses natural resources. The topic of continual decreasing of emissions and making more environmentally friendly choices is a very recent one. In 2012 the Paris Agreement was conducted and until this year, 2017, it has been signed by 195 independent countries and European Union. The deal commits the signed parties to lower their CO_2 emission by the year 2020 (Bjornali, Ellingsen, 2014). This has led to an increase of government interference on the cleantech markets and also the change in their donation policies (Liobikiene, Butkus, 2017). This topic is being very much discussed at the time of writing this paper, because of the withdrawal of the United States of America from this deal signed in May 2017 by President Trump, which brings great uncertainty to the expert audience as well as the wider public around the world.

Gaddy et al. (2017) proposed a paper on the development of venture capital investments in the cleantech sector. There was a significantly increased investor appetite in the sector that proceeded to exist until its peak in 2008 when it was sharply ended by the financial crisis of that year. In the period from 2010 to 2012 venture capital reverted back to said sector.

However ongoing results from this period showed that investments made in the cleantech sector were performing poorly in comparison to other sectors. These investments were more likely to fail, and if they did succeed, the return on investment was lower compared to other sectors. Starting in 2012 the venture capitalists began to fall back to sectors with more promising statistics on success and profitability of the investment. The 2017 study also suggests that the Paris Agreement will affect this situation, nevertheless the setback in investments continues to exist in the sector. That said, there is a very recent problem of cleantech startup funding, which is very likely to be caused by the significant risk that is inseparable from this particular sector (Huang, 2015). The high-tech sector as a whole is characterized by the development of new technologies that are expensive and difficult to predict with respect to usability and feasibility.

Therefore this paper focuses on the business failure prediction in the specific area of cleantech startups. Business failure prediction is an often researched topic for its numerous application and the influence on global markets, such as the paper proposed by Janáková (2015) dealing with success prediction of technology start-ups on Slovak market.

Some recent studies, such as Wang and Wu (2017) or Xu et al. (2014), suggest soft computing methods for this purpose, specifically fuzzy logic and its variations. The existing studies, however, are concerning companies with longer history such as well established businesses. Companies with a very short history are the concern of this paper. 17 variables were identified to describe the overall health of a startup with a dependent variable that represents the success or failure of each entered company after the first five years of its existence. Due to the nature of the examined sector there is no possibility to make use of conventional statistical methods of analysis and therefore the fuzzy approach was chosen. Fuzzy reasoning offers better possibilities to study the said sector, even with a limited number of companies to examine.

1. Fuzzy reasoning

1.1. Research background

Fuzzy logic is a component of soft computing methods along with rough logic, machine learning and evolutionary computing. Compared to a traditional set, a rough set consists of elements of a non-binary relationship between objects in the set (Kim et al., 2017). Veselý, Klockner and Dohnal (2016) published a study on the comparison of a linear regression model and a fuzzy logic model when used as a method of prediction of recycling behaviour. The conclusion of this study suggests that fuzzy logic is better in predicting this behaviour than linear regression and it can also perform analysis on predictors that are only vaguely known.

In the area of valuing a business, the expert knowledge describing the examined company is often described in imprecise words, often called “fuzzy” words. Fuzzy set theory was introduced in 1965 by Lofti A. Zadeh (1965). These sets allow data to be evaluated that are described in qualitative terms and to interpret them in a comprehensible and intuitive manner.

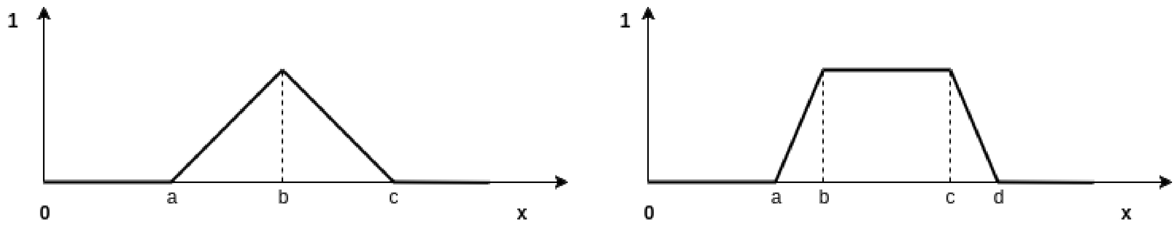
1.2. Fuzzy computation and modelling

A fuzzy set is distinguished from a crisp set by the grades of membership of each fuzzy value (Zadeh, 1965). A fuzzy set A is defined in space of discourse, X is a set of pairs:

$$A = \{(x, \mu A(x)), x \in X\}, \forall x \in X \quad (1)$$

The member function that characterizes the set A is μA : X → [0; 1], in other words the membership of each element can be of any value from 0 to 1. In summary, the fuzzy logic uses if-then inference rules to reach conclusions about the examined data using fuzzy sets as the objects of reasoning. The fundamental logic used in the fuzzy analysis is that by the comparison of the similar units, predictions about characteristics and contextual information that were previously unknown, can be made. The fuzzy model can be summarized in the five following steps: Empirical input → Fuzzification → Inference mechanism → Defuzzification → Crisp output (prediction). Empirical input consists of any data that are relevant to the overall prediction model. This input often comprises data of inconsistency units. In our examined case it is typically millions of dollars, percentage of interest rates or of surveyed subjects, and years. An important term to define is a fuzzy set, which is a collection of elements that each belong to a set with a degree of membership described by a membership function (Veselý, Klockner and Dohnal 2016). The membership function is defined by a curve which describes how each membership function fits the expert opinion. The proximity of the value μA(x) to 1 determines how x belongs to set A. There are several membership functions commonly used (Dubois and Prade, 1978), triangular, trapezoid, linear, sigmoidal, π-type, and Gaussian. For the purposes of this paper the triangular and trapezoidal membership functions have been used (Zavadskas et al., 2017).

The triangular function is characterized by three values, whereas the trapezoid is characterized by four, as seen in Figure 1 where the ***a***, ***b***, ***c*** and ***d*** points are representing the grade of membership by its proximity to 1. Alternatively it can be proclaimed that all values of the measured variable belonging to the interval between ***a*** and ***d*** belong to a fuzzy set to a non-zero degree.



Source: Zavadskas et al., 2017

Figure 1 Triangular and trapezoidal fuzzy membership functions

1.3. Fuzzification, knowledge base and defuzzification

Fuzzification is a process of changing the empirical input into fuzzified values, which are simplified into qualitative values. Each qualitative value, otherwise called a variable, is then represented by the above mentioned points a, b, c and d. All the variables are assigned a dictionary describing the ranges of its values. The dictionary of our examined model is shown in the chapter below in Table 1.

The fuzzified input returns a so-called knowledge base which consists of conditional statements. Each of the described cases - in our case each of the examined startup businesses - is represented by one conditional statement R. The statement R is equal to if-then statements, i.e if A and B then C. After all of the knowledge bases have been established, the inference mechanism processes this base and allows the prediction of unknown variables for new cases (derived from the work of Mamdani and Assilian, 1975). A final step is using a "centre of gravity algorithm" which returns crisp numerical predictions. While there are naturally other methods to achieve this, the centre of gravity algorithm proposed by Patel and Mohan (2002) is probably the most commonly used one.

It is important to note that very often it is not the case that there exists a perfect solution to a fuzzy logic model. The model and its parameters can be altered if the model's performance shows to be poor, as has been observed for example by Chand and Wang (1996), and Fisher (2006).

2. Experimental results and analysis

2.1. Data gathering and fuzzification

Data on cleantech startups and and performance after five years of existence were gathered and used as an input into the fuzzy set model analysis.

At the beginning of this study, two analytical approaches were considered, one using rough sets and the other using fuzzy sets. As both are very similar to each other, the process of the analysis was almost identical. Two experimental tests were conducted, both using the same data structure and the same model, but using a different methodology. The tests, conducted using rough sets, proved it to be an unfit method for analyzing the examined data set as it is not large; the output results with rough sets proved to be too vague and indefinite. Therefore the used method for business failure prediction is fuzzy sets, which showed far more promising results after the initial testing of the proposed model and was favoured to be the main experimental method for the rest of the study.

2.2 Fuzzy logic model

There are 17 attributes that can each have up to five values. As all the raw data collected on the examined startups contained many different units of measurement and varied in many different aspects, fuzzification of crisp data was performed in order to reduce the options, each attribute having up to five values at the most. With respect to the chosen data structure, the examined startups have been described as precisely as possible. 13 cleantech startups with a very short history were taken up as a part of this study and the data on the above mentioned attributes have been gathered and simplified into the qualitative form seen in Table 1. Each of the valued startups have been given a coefficient of their significance to the overall model, as not all of the examined attributes could have been measured precisely due to the short existence of the firms.

Each of the options per attribute then has to be given a fuzzy membership function. As mentioned above the fuzzy membership functions used in this paper were the trapezoidal and triangular functions. This can be seen in the short example in Table 1, where the column "type" determines the membership function to be either 7 - triangular or 8 - trapezoidal. Points 1 to 4 are then representing the values of ***a*** to ***d*** in the membership function.

Table 1 Model dictionary example

Attribute / Variables		Function type	a	b	c	d
CEO experience	CEO 4					
very experienced	VEX	7	5	7	9	
experienced	EXP	7	3	5	7	
limited experience	LEX	7	1	3	5	
not experienced	NEX	7	0	1	3	
Startup valuation	SVA 5					
very high	VHI	8	6	8	10	999
high	HIG	7	3	5	7	
medium	MED	7	1.5	3	5	
low	LOW	7	0.7	1.5	2	
very low	VLO	8	0	0.3	0.7	1
Business stage	BUS 5					
mature	MAT	8	0.6	0.75	0.85	1
sustainable	SYS	7	0.45	0.6	0.75	
growing / surviving	GRO	7	0.3	0.4	0.5	
early existence	EEX	7	0.2	0.3	0.4	
conceptual	CON	8	0	0.1	0.2	0.3
Market size	MSI 3					
large	LAR	8	17	37	57	999
medium	MED	8	3	10	17	24

small	SMA	8	0	1	2	3
Investor appetite	IAS 5					
very high	VHI	8	15	30	50	999
high	HIG	8	4.8	7.5	15	22
medium	MED	8	0.8	2.2	4.8	6.2
low	LOW	8	0.4	0.6	0.8	1.2
very low	VLO	8	0	0.2	0.4	0.6

Source: own work

2.3 Data input

Each of the examined startups, labelled as cases, describe closely the situation in a startup with a short history. There is a short sample of parts of these cases in Table 2 below "st1-5" represents 5 start up firms and "v1-6" the first 6 variables gathered, "c" is a coefficient of the significance of each case to the overall prediction model.

Table 2 Input values into the fuzzy model

	v1	v2	v3	v4	v5	v6	c
st1	experienced	experienced	multiple areas	medium	growing / surviving	local	1.0
st2	experienced	experienced	one-sided	high	early existence	none / new brand	0.7
st3	limited experience	very experienced	broad	low	early existence	none / new brand	0.8
st4	experienced	limited experience	broad	medium	conceptual	none / new brand	0.6
st5	very experienced	very experienced	broad	high	sustainable	none / new brand	0.9

3.4 Data consistency testing

After establishing the fuzzy model as shown in the previous chapter it is possible to begin the fuzzy dialogue that is the desired output of the process as a whole. However, as mentioned in the chapters above, the fuzzy model and especially its parameters often need to be altered and adapted after some previous tests. The main goal of this paper was to perform testing of the produced model to create a consistent fuzzy model that would offer reliable output, in our case a viable prediction of success or failure of examined startup.

For the consistency testing a fuzzy computer software was used. As an input to the analysis software the dictionaries belonging to all the used variables and all of the studied cases of startups were used. An example of dictionaries for the variables "Startup valuation" and "Market size" can be seen below in Table 3, examples of studied cases are displayed in Table 4.

Table 3 Example of variable dictionary input

```
SVA 5
VHI 8 6 8 10 999
HIG 7 3 5 7
MED 7 1.5 3 5
LOW 7 0.7 1.5 2
VLO 8 0 0.3 0.7 1.5
MSI 3
LAR 8 17 37 57 999
MED 8 3 10 17 24
SMA 8 0 1 2
```

Table 4 Example of cases input

EXP EXP MUA MED GRO LOC NON NEU LOC LAR NEW LUX LOW COT AVE LOW FGR 1.0
EXP EXP ONS HIG EEX NON LOC NEU SMA MED SAT NOR HIG PEA AVE INL STG 0.7
LEX VEX BRO LOW EEX NON NON POS LOC SMA SAT LUX VLO EXP LOW INL FGR 0.8
EXP LEX BRO MED CON NON LOC NEU GLO LAR SAT NOR MED PEA AVE INL FAI 0.6
VEX VEX BRO HIG SUS NON LOC POS LOC LAR SAT LUX VHI PEA AVE DEF FGR 0.9
EXP EXP BRO VHI MAT GLO GLO POS GLO MED NEW NEC HIG PEA AVE LOW FGR 1
VEX EXP MUA VHI SUS GLO NON NEU GLO LAR NEW NOR HIG EXP AVE INL STG 0.8

The principle of the analysis of the above mentioned and displayed data is comparing all variables in the cases and evaluating its impact on the result variable, which in this case is the success or failure of the startup after five years.

Several consistency tests have been conducted and by each iteration the results were significantly improved. The consistency and overall quality of the model is measured by the strength of similarity of cases and their influence on the dependant variable. Given the fact that there are more examined variables than there are studied cases it cannot be expected that the similarity of the cases will be very strong. After conducting the first experimental test, the best similarity of cases was 3% while the last testing showed several similarities stronger than 20%. Considering the conditions of this proposed data set, this is considered as a consistent enough fuzzy model to be used on examined data. In the following Table 5, the results of the final consistency test are shown, the most important to note are the last three numbers in each row that represent the case that is similar to currently analyzed case, the percentage of similarity and lastly the variable that causes the percentage not to be higher. These results allow alterations to be made to the model to further improve its consistency and reliability.

Table 5 Example of model consistency testing result

1 LOW 8:0.00 1.00 1.50 2.00 ǀ 0.038 9 0.03 13 11 0.24 13
2 LOW 8:0.00 1.00 1.50 2.00 ǀ 0.088 9 0.09 9
INL 8:1.50 3.00 4.50 6.00 ǀ 0.079 10 0.18 11
9 LOW 8:0.00 1.00 1.50 2.00 ǀ0.115 1 0.03 13 11 0.12 15
INL 8: 1.50 3.00 4.50 6.00 ǀ 0.088 2 0.21 9 11
FGR 8: 25.00 50.00 75.00 100.00 ǀ 0.038 1 0.04 13
FAI 8: 0.00 20.00 25.00 30.00 ǀ 0.099 10 0.10 11

Discussion

This study dealt with finding an appropriate tool for the success prediction of startup firm in a specific area. The prime limitation of the studied sector was the small number of studied subjects, which does not allow for using conventional methods such as statistical approach. This paper revealed fuzzy method to be an effective approach for research studies dealing with prediction based on a small number of known variables.

Conclusion

This paper has proposed the soft computing approach as a method to be used to predict business failure on the specific business sector of cleantech startups. Two components of soft computing methods were considered in the pre-research phase of the study, fuzzy and rough sets. The pre-research phase showed that fuzzy sets are best fitted for the examination of studied data sets, mainly for the reason of the limited number of studied subjects. Several cleantech startup firms in their early years of existence have been analysed and their data structured into a set input and the computation conducted with fuzzy analysis software.

The intended purpose of the created fuzzy model was to discover underlying connections between the business' elements and its failure or success after five years on the market, which is generally believed to be the most crucial period of a startup's life. Elements of examined startups were analysed, structured and simplified to fit into the fuzzy logic model.

The consistency of the fuzzy logic model was the primary goal of this paper as it effectively determines the model's viability, in other words, its ability to make predictions about newly added cases. Several tests were conducted and by each iteration the input model was altered to improve its consistency. From the first iteration to the last, the similarity between cases was raised by nearly 20 %, from 3 % to 22.5%. Considering the conditions of this analysis, primarily the scarcity of relevant data input, the final similarity of 20% is considered to be sufficient and therefore the created fuzzy logic model is suitable to be used for intended startup success analyses.

References

Bjornali E. S., Ellingsen A. (2014). Factors Affecting the Development of Clean-tech Start-ups: A Literature Review. *Energy Procedia*, 58, 43-50. Doi: 10.1016/j.egypro.2014.10.407

Chang, N.-B., Wang, S. F. (1996). Managerial fuzzy optimal planning for solid waste management systems. *Journal of Environmental Engineering*, 122(7), 649-658. Doi: 10.1061/(ASCE)0733-9372(1996)122:7(649)

Dubois, D., Prade, H. (1978). Operations on fuzzy numbers. *International Journal of systems science*, 9(6), 613-626.

Fisher B. E. A. (2006). Fuzzy approaches to environmental decisions: application to air quality. *Environmental Science & Policy*, 9(1), 22-31. Doi: 10.1016/j.envsci.2005.08.006

Gaddy B. E., Sivaram V., Jones T. B., Wayman, L. (2017). Venture capital and cleantech: The wrong model for energy innovation. *Energy Policy*, 102, 385-395. Doi: 10.1016/j.enpol.2016.12.035

Huang, J. C. P. (2015). *Venture capital investment and trend in clean technologies*. Springer New York. Doi: 10.1007/978-1-4614-6431-0_11-2

Janakova, H. (2015). The success prediction of the technological start–up projects in Slovak conditions. *Procedia Economics and Finance*, 34, 73-80. Doi: 10.1016/S2212-5671(15)01603-2

Kim, Y., Ahn, W., Oh, K. J., Enke, D. (2017). An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms. *Applied Soft Computing*, 55, 127-140. Doi: 10.1016/j.asoc.2017.02.006

Liobikiene, G., Butkus, M. (2017). The European Union possibilities to achieve targets of Europe 2020 and Paris agreement climate policy. *Renewable Energy*, 106, 298-309. Doi: 10.1016/j.renene.2017.01.036

Mamdani, E. H., Assilian, S. (1975). An experiment linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1), 1-13. Doi: 10.1016/S0020-7373(75)80002-2

Patel, A. V., Mohan, B.M. (2002). Some numerical aspects of center of area defuzzification method. *Fuzzy Sets and Systems*, 132(3), 401-409. Doi: 10.1016/S0165-0114(02)00107-0

Veselý, Š., Klockner, C. A., Dohnal, M. (2016). Predicting recycling behaviour: Comparison of a linear regression model and a fuzzy logic model. *Waste Management*. 49, 530-536. Doi: 10.1016/j.wasman.2015.12.025

Wang, L., Wu, C. (2017). Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map. *Knowledge-Based Systems*, 121(1), 99-110. Doi: 10.1016/j.knosys.2017.01.016

Xu, W., Xiao, Z., Dang, X., Yang, D., Yang, X. (2014). Financial ratio selection for business failure prediction using soft set theory. *Knowledge-Based Systems*, 63, 59-67. Doi: 10.1016/j.knosys.2014.03.007

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.

Zavadskas, E., Turskis, Z., Vilutiene, T., Lepkova N. (2017). Integrated group fuzzy multi-criteria model: Case of facilities management strategy selection. *Expert Systems with Applications*, 82, 317-331. Doi: 10.1016/j.eswa.2017.03.072