“PAISII HILENDARSKI” UNIVERSITY OF PLOVDIV
FACULTY OF ECONOMICS AND SOCIAL SCIENCES
DEPARTMENT OF FINANCE AND ACCOUNTING

GEORGI PETROV KALOFEROV

SUCCESS FACTORS OF START-UP COMPANIES

ABSTRACT

of dissertation thesis for obtaining
the educational and scientific degree “doctor”
by field of higher education *3. Social, economic and legal sciences*
in professional field *3.8. Economics*
doctoral program *„Finance and Accounting“*

Supervisor:
Professor Stanimir Kabaivanov, Ph.D.

Plovdiv
2023

The dissertation thesis was discussed at the Department Council (DC) on 24.02.2023 and is scheduled for defence by a decision of the DC of the Department of Finance at the Faculty of Economic and Social Sciences of Paisii Hilendarski University of Plovdiv.

This dissertation paper contains an introduction, three chapters, a conclusion, a bibliographical reference, a list of abbreviations used, a list of tables (52 tables), a list of figures (35 figures) and appendices (3). It consists of a total of 201 pages, of which: an introduction, three chapters and a conclusion – 153 pages, appendices – 32 pages, bibliography of sources – 16 pages. The bibliography includes 159 sources.

The publications about the dissertation work are two independent and one co-authored.

The public defence of the dissertation work will take place on **02.06.2023** at **13.00h** in **room 126**, Rectorate of Paisii Hilendarski University of Plovdiv at an open meeting of the Scientific Jury, appointed by Order ………….. by the Rector of Paisii Hilendarski University of Plovdiv.

The author is a full-time doctoral student in the Department of Finance of the Faculty of Economic and Social Sciences at Paisii Hilendarski University.

The materials on the defence are publicly available on the website of Paisii Hilendarski University of Plovdiv (https://uni-plovdiv.bg).

I. OVERVIEW OF THE DISSERTATION THESIS

1. Relevance and Significance of the Dissertation Subject

The relevance and significance of dissertation subject is determined by the increasing need for financing of new start-ups. Although risk financing historically originated in the USA and is currently most developed in this region, with the processes of financial integration, its spread is observed in other countries as well. The topic is increasingly relevant in Bulgaria, where in recent years a few venture capital funds have been created, and which already have investment results. Considering the importance of venture capital for the development of new products and ideas in areas of the economy with high added value – such as technological, biotechnological, energy sources, the role of venture capital funds in the financial support of start-up companies will be increasingly relevant. Although a very small share of start-ups is venture capital-funded, an extremely large proportion of companies that go public through an initial public offering have received venture funding at some stage of their development (Kaplan and Lerner, 2010). Therefore, once the factors that lead to the success of a startup are derived and analyzed, they can be used as investment criteria when choosing whether to invest in a startup. And vice versa - using the investment criteria of venture capital managers, it is possible to predict which enterprise would be successful and reach an Initial Public Offering (IPO). This applies equally to the topic of entrepreneurship such as business processes and logic, business strategies, product life cycle. (Drucker, 2014; Cunningham and Lischeron, 1991; Kirzner, 1973; King and Levine, 1993; Van de Ven et al., 1984; Deakins and Freel, 2002; Bygrave and Hofer, 1991; Kent et al., 1982). A similar statement could also be made regarding the topic of the emergence, development, and contribution of venture capital managers in the process of creating innovations (Kortum and Lerner, 1998; Kortum and Lerner, 2000; Florida and Kenney, 1988; Lerner and Gompers, 2001; Kaplan and Lerner, 2010; Gompers, 1994). The relationship between the two is a topic that needs further research. Examining the regional concentration of venture capital funds, it can be found that these are the places with the most emerging innovative products - the USA (Silicon Valley), Europe (London, Barcelona, etc.) and others. Conversely, in countries where there is no or a negligible percentage of venture-financed enterprises, there is also a lack of innovation. For this reason, the values of the indices that measure the technological and innovation development of the countries have low levels. Therefore, establishing the factors that lead to success in business ventures will be very important especially if they are connected in one way or another with venture capital managers, and the topic of attracting and creating such funds will become more and more urgent, and perceived by less innovative countries.

2. Object of the study

The object of study in the dissertation are the start-up enterprises created after 2010.

3. Subject of the study

The subject of research is the activity of start-up enterprises in various directions - attracting risk financing, their investment activity, the industry in which they operate, their participation in public life and the media, patent activity and other characteristics implying the ultimate success of the created products and services.

4. Main purpose of the study

The main purpose of the study is to distinguish the factors that most strongly influence the probability of success of start-ups.

5. Sub-goals and tasks

5.1. Sub-goals

Within the scope of the study, the main purpose is decomposed into two sub-goals:

- Identifying the significant factors that influence the success of start-ups.
- Choosing an appropriate machine learning algorithm that predicts with high accuracy (over 85%) the success of start-ups. This will protect investors from investing in failed businesses.

5.2. Tasks

To achieve the set goals, the following research tasks have been formulated:

1) Analysis of the scientific literature and discovery of key factors for the success of start-ups.

2) To discover the most frequently used investment criteria of different types of investors.

3) To present the essence of venture financing as a success factor.

4) To derive the appropriate characteristics to measure the potential for success of the start-up.

5) To apply advanced machine learning models to unexplored startup data.

6) To test the reliability of the Crunchbase.com database as a source of information for startups.

7) To select appropriate models that have the highest accuracy for forecasting and apply them through a concrete example with a financial result for the investor.

6. Research thesis and hypotheses

The research thesis is that traditional financial models for company valuation cannot be applied in the context of start-up companies and cannot be used to evaluate their success due to the number of assumptions that are associated with them. Also, traditional approaches do not imply returning information about the factors that determine the success of start-ups.

The proposed research thesis gives grounds for formulating the two research hypotheses that complement it:

Hypothesis 1: The combination of financial and non-financial variables increases the accuracy of models for predicting the success of start-ups.

Hypothesis 2: Machine learning methods can help obtain more accurate estimates of the probability of success of start-ups compared to traditional financial analysis.

7. Limitations in the Scope of the Study

Limitations in the scope of the study are primarily related to the empirical data used.

The analysis does not include the following data:

- The received funding from relatives, friends, incubators, accelerators, and other sources that are small in volume, but important for the initial stage of development of start-ups.
- The lack of obligation of start-ups to disclose public information.
- The lack of inside information about start-ups that venture capital funds, business angels and other types of investors have.
- The geographical limitation of the study, due to the lack of data on start-ups and venture capital funds from all countries.

8. Used analytical tools and approaches

To verify the research hypotheses, as well as to achieve the goals formulated in the dissertation, an empirical study was conducted, which included the following classification techniques:

- Logistic regression
- Support Vector Machines
- Adaptive Boosting – AdaBoost
- Gradient Boosting
- Light GBM
- Decision Trees
- XG Boost
- Random Forest
- Extra Trees
- Linear Discriminant Analysis
- K-nearest neighbors' algorithm
- Multi-layer perceptron

II. STRUCTURE AND CONTENT OF THE DISSERTATION RESEARCH

This dissertation contains an introduction, three chapters, a conclusion, appendices, a bibliography, a list of abbreviations used, a list of tables, and a list of figures. The text is structured as follows:

Introduction

1. Relevance and Significance of the Dissertation Subject
2. Applicability of expected results
3. Object of the study
4. Subject of the study
5. Main purpose of the study
6. Sub-goals and tasks
   6.1. Sub-goals
   6.2. Tasks
7. Research thesis and hypotheses
8. Limitations in the Scope of the Study
9. Used analytical tools and approaches
10. Expected novelty, effect of the study, applicability of the results

First Chapter

1.1. Terminological definition of the main concepts

1.2. Description of the current startup ecosystem

2. Startup valuation methods

2.1. Modern methods for valuation of start-up companies
2.2. Traditional valuation methods
2.3. Alternative valuation methods
2.4. Application of traditional and alternative methods for valuation of start-ups
2.5. Heuristic rules and criteria of venture capital funds
2.6. Summary of traditional, modern, and alternative methods of valuation
3. Contemporary research on the factors determining the probability of success of start-ups
3.1. Factors influencing the success of start-up companies
3.2. Technology and innovation as a predictor of success
3.3. Impact of venture financing on initial public offerings
3.4. Impact of patents and intellectual property on the development of start-ups
3.5. Company founders as the main driver of success
3.6. Technology and social media as a tool for predicting the success of start-ups
3.7. Financial reporting of companies as a tool for predicting their success or failure
3.8. Applicability of machine learning in predicting the success of start-ups
3.9. A summary analysis of current research on start-up success factors
Second Chapter
A model for estimating the probabilities of success of start-ups
1. Description of the data
2. Description of variables
3. Choice of model input parameters
4. Description of data set balancing models
4.1. Mathematical formulation of balancing methods
4.2. Parameterization of SMOTENC
5. Data scaling methods
5.1. Mathematical formulation of the Standard Scaler
5.2. Standard Scaler parameterization
6. Methods for evaluating the qualities of the obtained models
6.1. Parameterization of Cross_value_score
7. Setting up the model
7.1. Parameterization of GridSearchCV
8. Description of start-up classification methods
8.1. Support Vector Machines - SVM)
8.2. Logistic Regression
8.3. Adaptive boosting (AdaBoost)
8.4. Gradient Boosting
8.5. Light GBM

8.6. Decision Tree
8.7. XG Boost
8.8. Random Forest
8.9. Extra trees
8.10. Linear Discriminant Analysis
8.11. K-nearest neighbors algorithm
8.12. Multi-layer perceptron
9. Indicators for evaluating the performance of models
10. Indicators of the importance of success factors
Chapter three Results and Discussion
1. Main characteristics of the population / input data used in the analysis
2. Results from an application of machine learning models
2.1. Confusion matrix, measures of accuracy, precision, recall, ROC AUC
2.2. Additional indicators for measuring the performance of classification algorithms
2.3. Significance of success factors of start-ups
3. Discussion of results
4. Conclusion
5. Recommendations and application of research results
6. Opportunities for future research
APPENDICES
Appendix 1: Parameterization of models for estimating the success of start-ups
Appendix 2: Descriptive Statistics Results
Appendix 3: Results of application of machine learning methods
Bibliography
Tables
Figures
Glossary of abbreviations used

III. BRIEF DESCRIPTION OF THE DISSERTATION STUDY

Introduction

The significance and relevance of the problem are offered in the introduction of the dissertation, the object and subject of research are indicated, the research thesis, hypotheses, purposes and tasks, the used analytical tools and approaches and existing limitations, as well as the structure of the dissertation research are formulated.

FIRST CHAPTER
THE VALUATION OF THE START-UP COMPANIES – A CHALLENGE FOR THE FINANCIAL SCIENCE

The first chapter of the dissertation is devoted to the introduction to the subject of the assessment of start-up companies and the inability of traditional approaches to predict their success. First, basic concepts are defined, and then traditional, modern, and alternative valuation methods are presented. Next, current research on the factors that determine the likelihood of success is presented.

Definition of a start-up

A start-up company is an organization that is created with the goal of generating high growth and developing a product or service in a highly uncertain environment.

Definition of success and risk

The definition of startup success is what is considered the "rule of success" in entrepreneurial ecosystems around the world, namely that the result of the activity is either an initial public offering or a merger and acquisition. We may define success as the achievement of either of these two states. A definition of success is also compared to a definition of risk (success or failure), as failure to achieve an initial public offering or M&A goal means classifying the company as unsuccessful, although this does not always mean bankruptcy, closure of activity/closed company etc.

1. Startup valuation methods

When studying the success of a company, different techniques can be applied - by assessing its value and comparing it to some benchmark (most often used for established companies with sufficient accumulated financial information), by assessing the probability of bankruptcy or by assessing the probability of achieving a given important and significant event (e.g., an exit through an IPO or M&A). This is also the approach used in the dissertation research.

1.1. Modern methods of valuation of start-up companies

These techniques are used to evaluate young companies before receiving funding from business angels, venture capital managers, etc. For the most part, they are not complicated techniques, with which it is easy to arrive at a possible valuation of the enterprise, and based on this, the investor decides whether to invest in the company. As they are mainly used for companies in the early stages of development, they are based on a complex of factors and they include not only financial variables, but also those related to the team, product/service, competition, etc.

Berkus method

This method of valuing start-ups, developed by Dave Berkus (The Berkus Method – Valuing the Early Stage Investment, 2022), first appeared in the book by Amis and Stevenson (Amis & Stevenson, 2001) and is characterized by early-stage valuation, especially of technology companies. It is gaining ground as a method due to the impossibility of relying on the projected income and cash flows, which are the main variables in classical financial models. The method is applied by determining values in a given currency for the different types of risks accompanying a particular company. The main components are: (1) a good business idea; (2) prototype (reduction of technological risk); (3) the quality of the management team (reduction of execution risk); (4) Strategic partnerships (market risk reduction) and (5) product launch or sales achievement (production risk reduction).

Risk factor summation method

Risk factor summation method (Achimská, 2020) (Montani, Gervasio, & Pulcini, 2020) which implies the determination of the sector average pre-money valuation of a company and which is the company's base valuation. It is then adjusted for twelve key risk factors. Through this method, investors are forced to consider in their evaluation many factors that would otherwise be ignored.

The score is calculated by assigning a rating (from -2 times to + 2 times) to each of the risks, which means adding or removing a value from the original base value.

Scorecard method

This pre-money valuation method was created by Bill Payne (Payne, The Definitive Guide to Raising Money from Angels, 2006) (Payne, 2011) and is also called the Benchmark Method because it compares a target company to similar companies from same region. The algorithm is to first take the average valuation of recent financings of similar companies relative to the one under consideration. As a second step, like the two methods presented, this assessment is adjusted for given factors such as management team, opportunity size, product/service and technology, marketing and sales channels, competitive environment, etc.

1.2. Traditional valuation methods

The basic approach in finance to estimate the value of firms is by converting a future value (expected in time after a given period) into a present value. This method of valuation allows us to include several factors that influence our valuation and that reflect the risk associated with the investment. Additionally, we could also compare different investments/businesses by comparing current values. Therefore, the same approach could be taken in the context of start-ups, albeit with additional clarifications given the differences between young companies and established ones.

Discounted cash flow method (DCF)

The discounted cash flow method can be found in all Corporate Finance and Valuation books (Brealey, Myers, Allen, & Edmans, 2022) (Brealey, Myers, & Marcus, 2020) (Damodaran, 2011) and it is easily adapted (modified) depending on the problem at hand – for example modification for start-ups. In this part of the thesis, the basic form of the method is described.

To apply the method and derive the intrinsic value of the company, several input parameters are needed: (1) the cash flow; (2) expected cash flow growth; (3) the value of the company's financing and (4) the value of the project at the end of the period under consideration. The indicator most often used to express cash flow is free cash flow, which in its simplest form is the result of gross cash flow after taxes minus investments. Considering that the model needs data on future cash flows, small changes in the assumptions (inputs) can lead to large deviations in the result. From this point of view, different scenarios are also made - pessimistic, optimistic and baseline to cover different values of the input data.

The model is based on the concept of Net Present Value, which discounts the free cash flow with an appropriate discount rate:

$$NPV = \sum_{t=0}^{n} \frac{FCF_t}{(1+r)^t}, \quad (1)$$

where FCF is the free cash flow and r is the discount rate.

The number of cash flows depends on the type of company and its life cycle (whether it is a Startup or an established company), on the horizon of the investor, the sector, etc. Often the period is up to 10 years, while for startups it is shorter: up to 5 or 7 years.

To obtain the final valuation of the company, to the net present value of the cash flows should be added the value of the project at the end of the considered period (Terminal Value), which represents the total discounted net value of all future cash flows that will occur after the period under consideration.

$$Company\ value\ = \sum_{t=0}^{n} \frac{FCF_t}{(1+r)^t} + Terminal\ value, \qquad (2)$$

where FCF is the free cash flow, r – the discount rate, and Terminal value – the value of the project at the end of the considered horizon.

Venture capital method

This method is the prerogative of venture capital managers and shows their perspective on the companies they analyze before providing them with funding. Together with the analysis of similar companies (Comparables), this method is most often used by venture capital funds. Andrew Metrick and Ayako Yasuda's book (Metrick & Yasuda, 2011) describes four common features:

- Evaluation of the exit value (of the investment) – this is the expected value in case of a possible successful exit – initial public offering, acquisition/merger, or direct sale. The exit value is determined either by relative (using multiples that compare similar companies) or by absolute valuation (DCF).

- Target return – percentage by which the exit value is discounted. Relatively higher values are used due to the significantly greater risk of investment failure. This rate is different from the cost of venture capitalists because only the probability of success is considered. Based on the number of years to exit and the probability of success, the target return can also be calculated.

- Expected retention rate – shows the percentage of retention of the ratio between the current investment and the possible successful exit. Because with each subsequent round of financing, the share in the company of the investors will decrease, especially for investors from an earlier stage of financing if they do not refinance at a later stage.

- The investment proposal – the comparison between the costs of the investment and its benefits, i.e., the valuation at exit weighted by the expected retention rate and divided by the money growth. In the standard VC method, the venture capitalists' investment is equal to the funds invested, while the benefit is their share of the company.

“First Chicago” method

This method is traditionally used in the initial stages of company development and builds on the Venture capital method. It uses the same analysis techniques regarding the projection period of future cash flows and the calculation of net present value. The difference with the second one is that three additional cases/scenarios are considered, which have different weights in the final evaluation - "Success", "Failure" and "Survival" scenarios. The weight of each depends on the likelihood of each scenario occurring.

1.3. Alternative valuation methods

Comparable companies’ analysis

This method is an essential tool of investment bankers when valuing companies (both public and private) or a given unit of the company. The main idea behind comparable company analysis is to provide a market standard for analysts to use in M&A situations, Initial Public Offerings, restructuring, etc. The comparison uses various financial ratios such as EV/EBITDA and P/ E. There are five steps to implement the model:

Step 1: Selection of a set of comparable companies.

Step 2: Locate the required financial information.

Step 3: Calculation of financial ratios - the most used indicators are EBITDA, enterprise value, ratios such as profitability, return on equity (ROE) and others.

Step 4: Compare companies – place the company against comparable companies to see the relative performance of the former versus the latter.

Step 5: Determining the valuation – by using the average values of the given multiples such as EV/EBITDA of the comparable companies, a valuation is also derived for the company, also determining the scope and potential growth.

Precedent transaction analysis

Similar to the analysis of comparable companies, this method also uses financial ratios to indirectly estimate the value of a company. As with the analysis of comparable companies, here the most basic task is to find companies that are comparable to the company under consideration. What is different here is to find comparable companies that have been acquired in a more recent period - for example in the last two to three years, although in practice this period must be revised and older deals to be considered, even though market conditions have already changed by now and this will lead to a bias of the valuation (Rosenbaum & Pearl, 2020). The authors describe the algorithm of performing the analysis, which is like that of the analysis of comparable companies:

Step 1: Selection of the set of comparable M&A transactions.

Step 2: Locate the required financial information.

Step 3: Calculation of key indicators, multipliers, etc.

Step 4: Clear the list of comparable deals and choose those that come close to the company.

Step 5: Derive the value in an indirect way by using the multipliers of the list of comparable deals selected in step 4.

1.4. Application of traditional and alternative methods for evaluating start-ups

Damodaran (Damodaran, 2018) (Damodaran, 2009) describes the main circumstances accompanying startups as: lack of operational and financing history, with little or no revenue and often operating loss, dependent on equity financing, hard to survive (high percentage of bankrupt companies), multiple equity claims as well as illiquid investment compared to public companies. According to the author, all these circumstances surrounding start-ups lead to difficulties in their valuation using the standard financial approaches to the valuation of companies such as the discounted cash flow method by looking for the intrinsic value, as well as when using the relative methods mainly due to the missing history.

When assessing the value of a company, investors emphasize on the following components:

- current assets
- the value of future growth/growth assets
- the risk associated with the discount rate
- the valuation when the company will be already mature.

Each of these components can relate to the circumstances surrounding the difficult valuation of young companies. According to Damodaran (Damodaran, 2018), current assets are not a sufficient basis for valuation because they are small in value in most cases. The second component, however, is the one to focus on – growth assets. However, due to the lack of historical information, it would be difficult to predict future growth (be it revenue growth or earnings growth). Moreover, the rule that value is created when the return on capital is greater than the cost of capital would hardly apply to young companies and to predicting their future growth, given that the former often has a negative value. The use of discount rates (as a major component of almost all financial models) according to the author is challenging, due to the lack of market valuation of the securities of non-public companies, regression cannot be applied to estimate the beta coefficient of the model based on historical data, and the difficult diversification of firm-specific risk. The inclusion of the failure risk component to increase the discount rate is contrary to the assumptions of the DCF model for the going concern principle. Finally, but not least, the author adds the component that has the greatest weight in the overall evaluation of the company – the value of the project at the end of the considered period. This assessment, however, is also hampered by problems such as predicting when growth will become stable.

According to Damodaran (Damodaran, 2018), the presentation of relative valuations is no different and they also have problematic assumptions that lead to misevaluation of start-ups. The first assumption is that all financial ratios (multiples) are related (scaled) to some common measure such as profit, book value, revenue, etc., therefore all these "denominators" in young companies can be either negative, zero or with small values, which will lead to wrong conclusions. The second assumption is ultimately related to the need to use ratios to compare with other companies, as is standard practice for publicly traded companies. In this case, however, when it comes to start-ups, it would be difficult to compare companies that do not have market prices and other metrics. The third assumption is that the use of alternative measures of risk such as beta coefficient and standard deviation to measure ownership risk cannot be applied due to the impossibility of calculating these measures given the short history of companies.

The main model that is applied in practice when evaluating start-up companies is the Venture Capital Method and it is considered the main tool of venture capitalists. However, according to Damodaran, this method also has its drawbacks, primarily related to the "cat and mouse game" between entrepreneurs and venture capitalists. The former will try to increase projected earnings, while the latter will try to lower them to lower the valuation. The other problem that the author considers is related to the discount rate, which is equated with the target rate required by risk investors. However, it also includes the probability that the firm will not survive, which creates problems in that the future value must be based entirely on the value of equity, and that the discount rate will not change, even though if the firm grows and the probability of failure decreases.

The private transaction ratio approach, according to Damodaran, involves using data on similar firms. The data should contain information about how much was paid for them and these values should be scaled to some variable such as revenue or profits and a financial ratio should be obtained. The disadvantages of this approach, according to the author (as well as the reasons for not using traditional corporate finance models in the dissertation) are:

- The lack of a database on private companies and the transactions in which they are involved.
- The price paid may also include additional features beyond the company price.
- Differences in timing – unlike public companies, where their shares can be constantly bought and sold, i.e., their current price reflect current information, private transaction ratios can have large gap periods between trades.
- Differences in accounting policies, lack of revenue, profits, etc.
- If there is a database, it is likely that a large part of it will refer to the US.

According to Damodaran, the success of venture capitalists is not so much about their valuation skills as it is about choosing companies based on the product they offer, the quality of management, and so on. The author adds that a good practice in determining the discount rate is the part that deals with the probability of survival or failure. It is proposed to create a model for companies that have succeeded and those that have not, based on company-specific characteristics - cash, age, history of entrepreneurs, business field, etc.

1.5. Summary of traditional, modern, and alternative valuation methods

As a follow-up to the proposed traditional, modern, and alternative methods, Table 1 summarizes all their advantages and disadvantages.

Table 1. Summary of Traditional, Modern, and Alternative Valuation Methods

Method	Application	Disadvantage	Advantage
Berkus method	In the early stages of development; technology companies	It is not universal for different countries; It applies until before the revenue is received and the company scales up	It does not rely on projected revenues and cash flows; Emphasizes risk assessment of different types of variables (not just financial)
Risk factor summation method	In the early stages of development; technology companies	It is not universal for different countries; placing the same weight on different factors is not always appropriate	It does not rely on projected revenues and cash flows; Emphasizes risk assessment of different types of variables (not just financial)
Scorecard method	In the early stages of development	It requires comparison with other companies, as there may not be a similar one in the region	It emphasizes risk assessment of different types of variables (not just financial) by assigning different

			weights
Discounted cash flow method	Established companies; Companies with positive cash flow	It relies on cash flows, which young companies may lack; There are a lot of assumptions	Enables comparison between different companies; Easy to apply
Venture capital method	Start-ups; Growth companies;	Higher valuations relative to DCF; There are many assumptions	This method avoids some of the disadvantages of the discounted cash flow method; Specified to the company's stage of growth; Specified to the characteristics of start-ups
"First Chicago" method	In the early stages of development	There are many assumptions	It is based on the Method of Venture Capital Managers, but specifies three scenarios - Success, Failure and Survival.
Comparable companies analysis	Valuation of public or private companies or a division of a company; Upon initial public offering or Merger/Acquisition;	Comparable companies and financial information about them may not be found	Easy to apply using standard financial ratios;
Precedent transaction analysis	Valuation of public or private companies or a division of a company; Upon initial public offering or Merger/Acquisition;	Focuses only on previous M&A deals; It focuses on a period of standard 2-3 years, by which time the business environment may have changed; Data may be missing	Uses clear and understandable indicators; Based on actual transactions (i.e., no assumptions)

Source: Author's interpretation

- The model proposed in the dissertation will be consistent with the advantages of the proposed methods. The characteristics of the author's model are as follows:

- Lack of assumptions – the data is used as it is in the database.

- It's universal – it can be applied to startups at any of their stages.

- Supplements the financial criteria with additional ones, taking into account the risk for the other types of variables.

- Easy to interpret once the algorithms are trained.

- Adaptive method – when new information about companies is available, the model learns itself and starts to provide more and more relevant results.

- It can be used for comparison between companies, giving information about a specific success result, as well as what are the factors that influence it.

- Can use both historical data and current data.

- Overcomes the requirement for raw and hard-to-get cash flow data by using an indirect measure. Thus, it overcomes the shortcoming of most traditional methods of positive cash flow.

2. Modern research on the factors determining the chances of success of start-up companies

In this section, summarized data from all scientific articles is presented. Success factors, methods used, and target variables for success are summarized. Table 2 presents the factors that determine the success of start-ups.

Table 2. The factors that determine the success of start-ups

Factor type	Success factors in literature	Used in the dissertation
Financial	Funding from venture capital funds	Yes
Financial	Funding from Business Angels	Yes
Financial	Funding from Investment Banks	Yes
Financial	Funding through the various rounds	Yes
Financial	Crowdfunding	Yes
Financial	Capital adequacy	No
Financial	Good financial control	No
Financial	Accounting information	No
General	Location	No
General	Patents	Yes
General	Trademarks	Yes
Industry specific	% of growth of the industry	No
Industry specific	Type of industry	Yes
Economic specific	Macroeconomic Indicators	No
Company specific	Client relationship	No
Company specific	Company age	Yes
Company specific	Participation in the media	Yes
Company specific	The network of connections - with suppliers, partners, universities, etc.	No
Company specific	Production facilities	No
Company specific	Employees with technical experience	No
Company specific	Participation in an incubator	No
Company specific	Participation in an accelerator	No
Company specific	Innovation	No
Company specific	Presence in social networks	No
Company specific	The business idea	No
Company specific	Business size	No
Company specific	The use of consultants	No
Entrepreneur specific	Education of entrepreneurs	No

Entrepreneur specific	Work experience of entrepreneurs	No
Entrepreneur specific	Income before starting the business	No

Source: Author's interpretation

Table 3 presents the target variables describing the definition of success in the literature.

Table 3. Success variables in the literature

Definition of success	**Used in the dissertation**
Reaching a certain level of income	No
Obtaining funding	No
IPO	Yes
M&A	Yes
Subsequent rounds of funding	No
Survival	No
ROI	No

Source: Author's interpretation

Table 4 lists the methods that have been used in the contemporary literature to study the success of start-ups and these are those that build on traditional models.

Table 4. Methods of analysis used in modern literature

Analysis method	**Used in the dissertation**
Logistic regression	Yes
Hazard	No
Neural networks	Yes
SVM	Yes
Adaboost	Yes
Gradient Boosting	Yes
K-NN	Yes
Decision trees	Yes
Random Forest	Yes
Naïve Bayes	No
Bayes Network	No
Extra Trees	Yes
Adaboost	Yes
Ensemble	No
LightGBM	Yes

Source: Author's interpretation

The advantages of the analysis in the dissertation compared to existing research are the following:

- Dataset consisting of 100,000 companies (reduced to 70,000 after data cleaning).
- A unique set of variables not found in any other research paper - in number and variety.
- More analysis methods were used compared to other studies, even ones that have not been applied in the literature so far.
- Deriving success factors in addition to predicting success.
- Achieve better results in terms of predicting success.

SECOND CHAPTER
A MODEL FOR ASSESSING THE PROBABILITY OF SUCCESS OF START-UP ENTERPRISES

In the second chapter of the dissertation, the methodology of the scientific research is presented. The variables used are comprehensively described, as well as the data for each of them. Given good practices in data analysis, the second chapter also explores data balancing and scaling models. The context in which the methods described in the dissertation can be applied is different, regardless of the type of investor, his specialization, objective, time frame for exiting the investment, the stage of development of the company, the size of the investment, geographical preferences, liquidity preferences, risk profile, activity level, etc. All twelve algorithms that are offered as analysis tools in the dissertation have been carefully selected based on the author's personal experience in his practical data analysis work. Also, not a small part of the tools has been validated in other studies about the success of start-ups. There is also a small portion that is experimentally included to test how they will perform on startup data.

The chapter concludes with a description of the performance indicators of the research model.

The research model construction is conducted through several phases:

The first phase is mandatory, regardless of the chosen analysis approach, because the result depends on the input data and to avoid the "Garbage in, Garbage out" concept. Therefore, data cleaning is one of the most important steps in the analysis. Preparing such a large data set is not an easy task, but this is one of the advantages of the analysis over those described in the literature review. To date, the literature rarely mentions the data cleaning approach in models for forecasting the success of start-ups. Moreover, much more up-to-date data is used than those in the literature, where most of the papers used the freely downloadable Crunchbase dataset from 2017.

The second phase relates to the so-called engineering the characteristics and deriving the variables that will be used in the model to assess the success of start-ups. After applying the first step and cleaning the raw data, the variables should be in a format that will yield the best results from the machine learning algorithms. The application of financial and non-financial variables is the main driver of the analysis and the advantage over the range of assumptions that is used in traditional models.

The third phase is to prepare the model with the appropriate parameters to avoid the various problems that are connected to machine learning algorithms, namely overtraining or undertraining. Also, at this step the final models are considered.

The fourth and final phase is to apply the analysis and derive the success factors of the start-ups, as well as indicators to prove the advantage of the classification algorithms for machine learning over traditional financial models. The purpose of this step is to highlight the advantages over traditional and modern business analysis methods, which are based on a number of assumptions and hard-to-register data, when it comes to non-public companies.

1. Description of the data

Finding reliable information about private (non-public) companies is quite a difficult task given their non-disclosure obligation. However, there are good examples of databases from which this type of information can be extracted to be subsequently amenable to analysis. Platforms that offer information about startups are several, examples of which are Crunchbase, Pitchbook, Dealroom.co, etc. The data that is used in the dissertation is registered by the most famous platform for start-ups - Crunchbase. After applying a filter to screen out companies whose founding date is after 12/31/2009, the data set totals 72,328 companies, which is the final number of start-ups that are the subject of research.

Of these 72,328 companies, 6,612 have an exit-type event – i.e., either have been acquired or made an Initial Public Offering. This is the first class of data that will make up the so-called "positive data class". For the remaining 65,716 companies, there is no information on whether they were acquired or whether they have a IPO, which means that they will be in the second class of data and will constitute the so-called "negative data class".

2. Description of variables

The variables presented in Table 1 were implemented in the research model. It is important to clarify that they were not chosen simply randomly or only because there was data available for them in the database. On the contrary, based on a deep analysis, what are the most common reasons why companies fail (such as lack of money or spending it too quickly without the possibility of replenishing it, lack of innovation, wrong sector, wrong time to release the product on the market, etc.) and the variables that the author considers potentially would contribute to the company's success and, respectively, their absence would contribute to failure, are also shown. Given the objectives of the dissertation, greater importance is given to financial variables and therefore their number is greater than the others.

Table 5. Variables used in the models

Target variable	**Description of target variable**
HasExit	A variable that implies: if the value is "1", then the company has been acquired or gone public. If the value is "0", it means that the acquisition or initial public offering event has not yet occurred for the company. The variable is derived based on another variable "Exit Date" - if there was such a date, then the variable is assigned a value of 1 if there is no date - a value of 0.
Features	**Description of features**
NumberOfIndustries	Number of industries the company is involved in (e.g., Cloud computing, Advertising, Software, etc.)
MonthsSinceFoundation	Months since foundation (company life in months) – the variable is derived from another variable “Founded date” (the date the company was founded) and measures the difference between the founding date and the date when the data was extracted from the Crunchbase database.
MadeAcquisition	A variable that indicates whether the company has made acquisitions during its operating activities. It is assigned a value of 1 if it has made acquisitions and a value of 0 if not. The variable is derived based on another variable "Acquisition status" and if has a value of "Made Acquisitions" 1 is assigned and vice versa.
HasInvestorOtherExits	If the investor who invested in the given company has "exit" events in other companies, then a value of 1 is set, otherwise a value of 0 is set. The variable is based on another variable Number of Exits (number

	of exits), and if there is at least one exit, a value of 1 is also assigned to the variable.
HasInvestorOtherInvestments	If the investor who invested in the given company has investments in other companies, then a value of 1 is set, otherwise a value of 0 is set. The variable is based on another variable Number of Investments, and if there is at least 1 investment, a value of 1 is assigned.
HasInvestorOtherLeadInvestments	If the investor who invested in the given company has investments as a LEAD investor in other companies, then it is assigned a value of 1, otherwise it is assigned a value of 0. The variable is based on another variable Number of Lead Investments (number of investments as a lead investor), and if there is at least 1 lead investment, a value of 1 is set.
HasLeadershipHiring	If the start-up has had at least one leadership hiring (i.e. hired a senior executive outside the founders – such as VP, CEO, etc.), then variable is set to 1, otherwise it is assigned a value of 0. The variable is based on another variable Last Leadership Hiring Date, which indicates the date of the last hiring of a senior leader. If there is such a date, then such a person was hired in the company and this is denoted by a value of 1.
NumberOfLeadInvestors	Number of lead investors in the given company.
NumberOfInvestors	Number of investors in the given company.
NumberOfAcquisitions	Number of acquisitions the company has made.
HasFounders	Because there are missing data on the number of founders (and in general this is an important variable when studying startup success), a variable was added to give a value of 1 when the value of the variable Number of Founders is greater than 0, thereby "supporting" the rows of data where there is founder information.
HasLeadInvestor	A variable indicating whether a lead investor has invested in the company.
NumberOfArticles	Number of articles that refer to the given company.
NumberOfFundingRounds	Total number of funding rounds.
NumberOfEvents	Total number of events the organization attended.
PatentsGranted	The number of patents granted for the given company, according to IPqwery
TrademarksRegistered	The number of trademarks granted for the given company, according to IPqwery

Source: author's interpretation

Techniques for balancing the minority class (SMOTENC) and for scaling the data (Standard scaler) are applied.

3. Models Performance Indicators

In the dissertation, the following indicators were used to evaluate the performance of the models: ROC, Accuracy, Precision, Sensitivity, Fbeta, Jaccard coefficient, Log loss, Confusion matrix, Matthew's correlation coefficient, Balanced accuracy.

CHAPTER THREE
RESULTS AND DISCUSSION

In the third chapter, the results obtained from the empirical research are presented and analyzed from a theoretical and practical point of view.

1. Results from the application of machine learning models

In the dissertation, it is emphasized that the most used indicator is Accuracy, which shows the correctly classified observations to the total number of observations and the goal is to maximize it, which would also mean a better forecast. In the context of predicting the success of startups, if accuracy increases, it means more correctly classified classes (in this case, successful or failed companies), which is extremely important for investors. It is especially important that the data classes are balanced, otherwise the metric can be misleading because the model may have a high accuracy value but not perform well in predicting the minority class. This has been overcome after applying the SMOTENC balancing technique. Based on this indicator, the three best models are as follows:

- Extra trees – 0.861
- Random Forest – 0.858
- Decision trees – 0.815

Based on the accuracy metric, we can conclude that the leading prediction models are those based on decision trees. Despite the little importance they are given in investment and corporate finance textbooks, the research presents them as an equal tool to traditional techniques. Given the high probability of failure of start-ups, using a tool that has 86% accuracy (as Extreme Random Decision Trees) would, on the one hand, protect investors from failed deals and, on the other hand, show them the ones that would be successful.

Next in terms of performance is the K-nearest neighbors' method, followed by boosting algorithms and the multilayer perceptron. They are followed by support vector machines, logistic regression, and linear discriminant analysis. Some algorithms such as logistic regression and the support vector machines performed rather disappointingly given their central role in other studies described in Chapter Two. It's possible their bad performance is due to the inability to perform Grid Search CV optimization technique, but it would hardly have improved their results by much.

The next metric - ROC AUC - measures the ability of the model to distinguish between positive and negative observations at all possible classification thresholds (for all possible values of these thresholds). These thresholds are the decision limits – whether an observation is positive or negative (in binary classification problems). Or in other words the thresholds themselves are used to convert the probabilistic output of the model into a binary prediction. Accordingly, what is specific about ROC is precisely that it concerns all these possible thresholds, which means that ROC considers all possible trade-offs between TPR and FPR that can be obtained by changing the thresholds. In the context of predicting the success of startups, the ROC shows how well the model differentiates between successful and unsuccessful companies. A higher value means more accurate classification between successful and unsuccessful, meaning that it is important for investors that the models they work with have a high ROC value. For example, if we increase the ROC of the model by 1 point, it means that the TPR of the model (sensitivity) of the model increases, while the FPR (1-specificity) of the model decreases.

Specifically, according to the model results, the three best performing models based on this metric are again those based on decision trees:

- Extra trees – 0.86
- Random Forest – 0.8582
- Decision trees – 0.8156

Although with a minimal difference between Extra trees and Random Forest, the former performs best, the difference in value between the three possibly being due to various factors such as the number of trees used in the model, the depth of the trees, the number of the features used, etc. However, regardless of which of the three models investors choose, they would contribute positively to the management of their startup portfolios.

The third indicator is the Precision-Recall curve (PR AUC). It shows us (like the ROC curve) a correlation between 2 indicators. In this case, these 2 indicators are precision and recall. As separate indicators, they will be considered below. When viewed as interrelated, they show the trade-off between precision and sensitivity. The need for this indicator is related to the fact that it is particularly useful in cases where the positive class is less common as it is in our cases, where successful companies are much less compared to unsuccessful ones. Therefore, this indicator is often more informative about model performance than accuracy or ROC AUC in similar cases.

In cases of predicting the success of start-ups, the PR AUC can provide useful information and guidance in identifying potential successful start-ups. E.g., with a high precision value, it means that the model makes a correct prediction of successful businesses, while a high sensitivity value means that the model identifies a large proportion of successful startups in the given data set. And finally, this indicator can also help in determining the optimal value of the decision-making threshold. The point on the curve where precision and sensitivity are highest represents the best trade-off between the two indicators and can be used as a threshold for making predictions. The three best performing models based on this indicator are the same as for the previous two indicators, namely:

- Extra trees – 0.8951
- Random Forest – 0.8918
- Decision trees – 0.8613

Given that extra trees have the highest value, this means that they best balance precision and sensitivity, meaning they would most accurately identify successful companies while minimizing the number of false positives. (The start-ups that were predicted to be successful but were not). The same conclusion can be applied to the other 2 models – random forest and decision trees, as their values do not differ much from the first one.

2. Discussion of the obtained results

The first part of the results discussion is related to the presentation of the different analysis algorithms and the obtained results related to accuracy, precision and other indicators. Based on the information from the classification reports, it can be seen that algorithms based on random decision trees such as Random Forest and Extra trees perform better than all other algorithms based on neural networks, boosting algorithms, etc. These results contrast with some of the analyses described in the literature review, such as Arroyo et al. (Arroyo, Corea, Jimenez-Diaz, & Recio-Garcia, 2019), where boosting algorithms perform best, while in ours they underperform those based on random trees. Apart from the cited study, the application of Extra trees is not found elsewhere, and here it is in first place in terms of presentation. The application of such an algorithm also makes sense in practice given its high success rate, speed, and efficiency in training and when applied to test data, as well as when setting parameters and hyperparameters. Most of the studies in the literature review emphasize the support vector machines (Böhm, et al., 2017), (Żbikowski & Antosiuk, 2021) and logistic regression, which in this study perform quite poorly and should rather not be applied in the study of the success factors of start-ups. Like Boosting algorithms such as Gradient Boosting, Adaptive Boosting, LightGBM and XGBoost have rather average performance in the analysis, although they are applied quite often in the literature. In terms of neural networks, as in the analysis of Hora et al. (Horak, Vrbka, & Suler, 2020) perform better than the support vector machines or the analysis of Ang et al. (Ang, Chia, & Saghafian, 2020), in which neural networks achieve a high classification result, but on the other hand, our analysis shows that they are outperformed by methods based on random decision trees.

The second and more important part of the analysis is related to the identified success factors. Although with minimal differences in significance between different algorithms, this point presents results (success factors) from the two best performing methods – Random Forest Algorithm and Extra Trees Algorithm.

The first factor in importance is the number of months since founding. This means that company age plays an important role in the success of startups. This result is similar to the analysis of Yin et al. (Yin, Li, & Wu, 2021) and Diaz-Santamaria and Bulchand-Gidumal (Díaz-Santamaría & Bulchand-Gidumal, 2021) who also found that age plays a role in determining success.

The second factor is related to the number of media articles about the given company. As can be seen from the results, it is important to "be in media" (technological, social, traditional), thereby increasing the probability of success of the Startup. This success factor is also part of the analysis by Scharliev et al. (Sharchilev, et al., 2018); Zhang et al. (Zhang, et al., 2017); Gloor and others. (Gloor, Dorsaz, Fuehres, & Vogel, 2013), which prove that media presence is of utmost importance for attracting both funding and achieving the cherished goal – IPO and M&A. This factor also includes the other similar, but not so important factor, namely - the number of events in which the company participates, and which are reflected in the database.

The third factor is the number of industries in which the company operates. This factor is also one of the investment criteria of venture capital managers, who are most often interested in the industry/sectors in which the company operates, and their potential. Through this proxy variable – the number of industries – the analysis attempts to infer whether involvement in more industries leads to success. The results of most algorithms show that the number of industries is one of the most important success factors. Most studies emphasize founders' experience in a given industry, but nowhere does the number of industries in which it operates appear as a factor. This is also one of the significant advantages of our research.

Number of funding rounds and number of investors/lead investors are the next important success factors for startups. This also proves one of the author's claims, namely that venture capital managers and other types of private equity investors are an important factor for the future success of companies given all the benefits they receive from partnering with these alternative capital providers. What is an advantage of the analysis in the dissertation is the use of additional variables for the number of investors and for lead investors. This is a contribution to the theory and opens the topic on the one hand about the syndicated approach to investing and its basis for future success, and on the other hand - about the importance of leading venture capitalists or other types of investors also participating in the syndicate of investors.

The next two success factors are related to patents and trademarks. As the analysis shows, they are an essential condition for the future success of start-ups. What is important to be mentioned is that the direct relationship between them and success is investigated in the research, and this is done for the first time in the literature, adding trademarks as well.

And as a final factor and signal of success is the number of acquisitions. This is also new to the literature, as this factor has not been considered before. As we can see, it is a particularly important signal for investors because if a company has been making acquisitions of other companies, it would mean that the company is doing well enough financially and generating enough cash flow.

To present the financial benefit of using such type of tools in a real environment, the success of the first Bulgarian unicorn Payhawk - a fintech company that offers a technological platform for combining credit cards, payments, expenses, etc., and is predicted to be acquired or go public given the recent capital raising events that attracted some of the leading venture capitalists. Currently, it is still a private company, which means it is a good example on which to test the models trained in the thesis and predict success.

It can be seen based on checking the Eleven Ventures' share in Payhawk Limited (which is the English company with 100% capital in the Bulgarian company) that Eleven Ventures (ELEVEN INVESTMENTS KDA) first received 309,046 preferred shares in the first stage of investment (pre-seed) and 143,600 in the next financing round (seed) or 452,646 total preferred shares (Payhawk, 2023). If we assume that in the event of reaching an initial public offering and converting the preferred shares to ordinary shares in a ratio of 1 to 1 (naturally, it matters a lot what is written in the preliminary agreements, as well as the specifics of the initial public offering, and accordingly the conversion can happen for example 4 ordinary for 1 preferred, but for the sake of the example we assume that the conversion is 1 preferred for 1 ordinary), it follows that as of today Eleven Ventures owns 3.58% of the company (452,646 shares out of a total of 12,658,916) . We assume that there will be no dilution of equity value to earlier investors in subsequent rounds of financing, as the company may have other rounds of financing. The company's final valuation upon receiving the last round of Series A funding is $1,000,000,000. Assuming the company sells now and that's the price that can be taken (which is likely to be quite a bit more when it comes to the IPO) or when it's time to exit the investment, then 3.58% of 1,000,000 000 billion dollars current valuation, it would return $35,800,000, which is more than 50% of the last fund raised (according to data from CBInsights in January 2022 the fund raised 67.92 million dollars) (CBInsights, 2023).

Table 6. Financial effect in case of a positive exit from the investment

309 046	Preferred shares in the pre-seed round
143 600	Preferred shares in the seed round
452 646	Total shares of Eleven Ventures as of January 2023
12 658 916	Total Shares of Payhawk Limited as at January 2023
3.58%	Eleven Ventures' stake in Payhawk Limited as at January 2023
$ 1 000 000 000	Evaluation from the last round of financing
1 000 000 000 * 0.0358 = = $ 35 800 000	Potential return for Eleven Ventures if, other things being equal, it exits the investment
67 920 000	The amount raised in the creation of Eleven Ventures' last venture capital fund
52.71%	The ratio between the potential return from Payhawk's exit and the amount raised by investors in Eleven Ventures' last fund

CONCLUSION

The dissertation presents the current situation with entrepreneurial activity and the key role of risk financing for the entrepreneurial ecosystem. The presented results confirm the increasingly important role of start-ups for the economy of a given region and for attracting capital. Defining success is inevitably an important step towards exploring the evolution of the ecosystem and what each company should strive for. Therefore, the main task is to derive what the success factors are and to create a model that can, based on these factors, predict with great accuracy whether a start-up will achieve success.

In a logical sequence, the main popular, traditional, and alternative methods of assessing value and success are presented, but their shortcomings make them difficult to apply to start-ups and to forecast their success. Among the shortcomings of the popular methods stand out their non-universality when applied in different parts of the world, impossibility to highlight which factors are important and which are of secondary importance, as well as which do not matter at all for the success of start-ups. They are often applied before funding is received, making them unsuitable for later stages of funding. Some of the models require finding comparable companies, which can be difficult.

With traditional methods, the disadvantages are even greater, as they mainly rely on cash flows, which in most start-ups are either absent or negative. Accounting information to derive cash flows is often difficult to find, or if it exists, access to it would be difficult and expensive. In addition, traditional methods based on discounted cash flows make many assumptions that can sometimes bias the valuation when examining start-ups at different stages of development – for example, allowing for over- or under-valuation when forecasting is wrong of future cash flows or wrongly chosen discount rate. In the alternative methods, the main disadvantages are related to finding comparable companies, M&A transactions, and financial information in general, which has become evidently difficult to find.

The dissertation also presents modern approaches and research for evaluating success factors in start-ups, in which the increasingly important role of machine learning algorithms and the use of variables other than the purely financial ones used in traditional ones are highlighted.

The author's research strategy presented in the dissertation overcomes all the assumptions of popular, traditional, and alternative methods by presenting a trained model that, using public data (without requiring a lengthy information retrieval process) for a given startup, can show with high accuracy the probability of the company's success (achieving the cherished goal of an Initial Public Offering or M&A). The model is also universal and can be applied to any company from any continent and industry. It supplements the financial criteria (which are often missing in start-ups) with other variables such as patent activity, media activity, as well as the role of venture capital funds. The output of the model is easy to interpret. The algorithm also adapts and self-learns as information about new or existing startups comes in. It can also be used to compare companies. In this way, the analysis of start-ups is easily automated, and time is saved for additional due diligence if necessary.

Within the individual chapters of the dissertation conclusions of varying degrees of generality are formulated. Some of them stem directly from the revealed problems. Another part refers to the results of the research and theoretical work carried out. A third is more in the nature of formulated proposals and guidelines for further research work in the area under consideration.

The research conducted on the success factors of start-ups allows the following more important conclusions to be formulated:

1. Traditional and alternative methods of valuation of start-up companies have many assumptions that practically exclude their application in an environment of lack of cash flows and accounting information, as well as of comparable companies. However, when it is decided to apply them, it happens in the later stages of development of companies, while modern methods are preferred in the evaluation of companies in the earlier stages of development.

2. Machine learning methods are a versatile tool that overcomes the assumptions of traditional and other modern techniques. They easily adapt to new information and once trained, they can easily be applied to evaluate companies, thus automating the evaluation process and at the same time the results they return are easy to interpret. In the analysis of the modern literature for evaluating the success of start-up companies, the use of these tools is mostly found at the expense of traditional ones. As input parameters, they are often trained not only with financial data (as with traditional ones), but also with company-, industry- and entrepreneur-specific data.

3. To be effective, not overtrained, and show solid prediction results, machine learning methods must be fed with cleaned, normalized, and balanced data. The application of scaling and balancing techniques has shown that this improves the prediction.

4. Algorithms based on random decision trees such as Random Forest and Extra trees show the best results in predicting the success of startups, followed by K-Nearest Neighbors and Neural Network algorithms.

5. The weakest results show the approaches most often used in the literature, such as Logistic Regression and the Method of Support Vectors, which methods are rejected when predicting the success of start-ups.

6. The Linear Discriminant Analysis model performs the weakest of all other models in terms of most of the indicators used, which should exclude its use for a similar type of classification task – identifying success factors of start-ups.

7. The derived groups of factors for the success of start-up enterprises are the following: (1) specific to the company such as age and the industries in which it operates; (2) financial, such as the number of financing rounds, the number of investors and lead investors; whether the company has acquired and (3) patent activity, which includes patents and trademarks.

IV. CONTRIBUTIONS OF THE DISSERTATION THESIS

Theoretical and Methodological Contributions

1. A critical comparative analysis of traditional, modern, and alternative methods of company valuation was made.

2. Modern approaches and studies for evaluating the success factors of start-ups are presented, through which the increasingly important role of machine learning algorithms and the use of variables other than the financial ones used in traditional approaches are highlighted.

3. The probability of success for new and start-up enterprises is defined in the conditions of limitations of financial and accounting information, which is a major obstacle to the application of traditional and popular methods of evaluating companies.

Theoretical and Applied Contributions

1. Three groups of success factors for start-ups are identified, which are not extensively examined in scientific research to date.

2. Alternative self-learning models for valuation of start-up companies using a set of financial and non-financial variables are proposed.

3. To present the financial benefit of using similar type of tools in a real environment, the success of the Bulgarian fintech company Payhawk, which offers a technological platform for combining credit cards, payments, and expenses, is predicted.

V. LIST OF PUBLICATIONS ON THE TOPIC OF THE DISSERTATION

1. Kaloferov, G. (2021). *Clustering European Venture capital funds based on investments*. *Vanguard Scientific Instruments in Management*, vol. 14, no. 1, 2021, ISSN 1314-0582. ISSN 1314-0582. pp 1-14

2. Kaloferov, G. (2022). *Predicting startup success using support vector machine*. AIP Conference Proceedings, Volume 2505, Issue 1. 2505, 020013-1 - 020013-6. https://doi.org/10.1063/5.0100684 (**Scopus**)

3. Kaloferov, G., & Kabaivanov, S. (2022). *Start-ups' success factors – Evidence from Balkans*. AIP Conference Proceedings. Volume 2449, Issue 1. 2449, 070008-1 - 070008-7. https://doi.org/10.1063/5.0090742 (**Scopus**)

DECLARATION OF ORIGINALITY

I declare that this dissertation is a complete author's product, and, in its development, no foreign publications and developments have been used in violation of their copyrights.