SCENTIA INTERNATIONAL ECONOMIC REVIEW Volume 1 - Issue 2 - 2022

Construction of a context-based system using similarity graph analytics on start-ups' critical success factors

Fabian Broszat, Comenius University, Faculty of Management, Bratislava, Slovakia Milan Fekete, Comenius University, Faculty of Management, Bratislava, Slovakia

Abstract

Current research in the field of data-driven analysis of start-ups' critical success factors is based on similar methods and not always differentiating aggregations of data. This paper examines the current data models used to analyse critical success factors and presents a new data model / analysis technique (algorithm) based on contexts with related examples.

Keywords: context-based critical success factors, graph databases, graph analytics, context-based analysis

JEL Codes: D01, D02, M13

1. Introduction

Current data models for the analysis of critical success factors (in particular start-ups are examined) are based on statistical methods and generalist aggregations of success factors that apply to various companies. The main problem of these general approaches is that they produce generalised rather than specific critical success factors that lead to non-specific results. Specific critical success factors are enormously important for predicting the success of start-ups and analysing their risks, as they can be used as a basis for making concrete decisions.

In the following paper, the current data models are therefore examined on the basis of their properties, then the current state of research is formed and a data model is developed and presented that incorporates context-based critical success factors (CCSF) [1] and demonstrates a new graph-based approach. This model is finally filled, executed, analysed, and discussed using sample data.

The title of this paper was chosen as a continuation of the paper "Why context matters for start-ups' critical success factors – the definition of context-based CSF" [1], as in this paper the CCSF presented is practically implemented and analysed by means of an experiment. Therefore, this paper aims at the application that is mapped via a context-based system.

2. Research questions

The following research questions are addressed in this paper. They are addressed according to the methodology in section 3.

- 1. Are graphs used in the current data models on critical success factors?
- 2. Are contexts collected in the data models or used to examine critical success factors?
- 3. How does a context-based data model look like using sample data and what insights / success factors can be extracted from it?

Research questions 1 and 2 are deliberately posed as closed questions in order to answer the current state of science at this point with a clear yes or no regarding the research question. Research question 3 is an open question because it can only be answered by the experiment carried out using the data model in this paper. More discussion is necessary here and therefore the question is to be marked as open.

3. Methodology

Research questions 1 and 2 are (deductively) negated on the basis of a literature review. For both research questions, it is relevant to outline the current state of research in order to design a novel data model to answer research question 3 and fill it based on sample data (inductively using the context-based critical success factors).

3.1 Theoretical methodology (related to research questions 1 and 2)

In the theory section for the analysis of current research, a literature review is mainly carried out. This is carried out systematically according to three criteria and at the end the current state of research is evaluated. The result of the systematic analysis then leads to the answer to research questions 1 and 2. The method is described in section 5 in more detail.

3.2 Practical methodology (related to research question 3)

In the practical part to answer research question 3, an experiment will be conducted using an IT-based setup with a self-developed algorithm and a self-designed architecture (plus script). This setup is filled with a data set and analyses are carried out on the basis of this data structure. These analyses should lead to the answer to research question 3. The methodology for conducting the experiment is described in section 6 in more detail.

4. Basic definitions and theoretical considerations

In the following, some basic definitions are outlined and explained. These definitions are fundamental for understanding this paper.

4.1 Start-up

There are various definitions of the term "start-up" and different views on this term. This paper defines the term according to [2]:

"A start-up is a human institution designed to create a new product or service under conditions of extreme uncertainty."

This definition is a more general one and thus also includes companies that have outgrown the initial phases. Since the transition from a start-up to an established company is usually fluid, a more general definition is advantageous in order not to restrict the selection of the objects examined in section 6 too much.

4.2 Context

There are various definitions to describe the term "context". A basic definition of the term "context" is the following:

"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and application, including the user and applications themselves." [3]

Based on this definition and further definitions the following summary can be stated:

"In summary, the context is an abstract model of the parameters that describe an entity. [...]" [1]

For more information on the term "context", please see [1].

4.3 Context-based critical success factor (CCSF)

The CCSF is described as follows:

"A CCSF is determined by context variables that make it work only exactly in its context. The CCSF is only valid there." [1]

In a nutshell, the difference between a general critical success factor (CSF) is that the CSF is valid in different contexts and is therefore generalistic. This leads to non-specific results as the following example shows:

"CCSF: 'We understand different technologies in the solar industry as well as cooling systems, and water treatment' \leftrightarrow CSF: 'Product Technology'" [1]

This CCSF is just valid for this specific context ("Technology for Solar energy") with the respecting parameters that belong to this start-up. The CSF "Product Technology" on the other hand, is generally valid and not context bound. Furthermore, the CCSF is more comprehensive and contains more information.

For more information on the CCSF, please see [1].

4.4 Graphs in general

In literature, a graph is defined as follows (no mathematical definition):

"Many real-world situations can conveniently be described by means of a diagram consisting of a set of points together with lines joining certain pairs of these points. For example, the points could represent people, with line joining pairs of friends; or the points might be communication centres, with lines representing communication links." [4]

For a better understanding of a graph, the following figure shows a basic example graph:

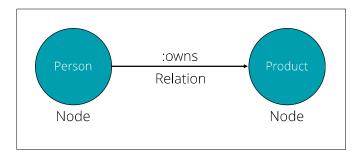


Fig. 1. Example graph (own creation)

From the presented definition and illustration, it can be summarised that 1-n data entities (nodes) are connected to each other via 1-n relations (edges). In addition to this, nodes and edges can also be named. This can look as follows for a large number of nodes and edges (larger graph):

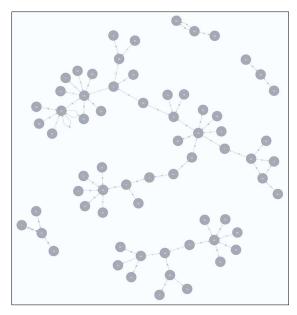


Fig. 2. Example graph 2 (schematic; own creation)

In this figure, it can be seen that nodes can have multiple edges and can also reference themselves, thus directing an edge towards themselves.

4.5 Graph databases

There are various databases on the market that can handle graph structures. In this paper, neo4j is used. Further technical explanations can be found on the manufacturer's website (this is not part of this paper).

4.6 Graph analyses

Based on this network data structure, analyses can be carried out and iterations can be made more quickly via the networking / relations of the data entities. Therefore, there are, for example, the following algorithms:

- *Pathfinding:* Pathfinding calculates the shortest path between two networked nodes.
- *Centrality:* Central nodes (nodes with the highest degree of interconnectedness) are identified in a cluster.
- *Community Detection:* This algorithm can be used to detect clusters and groups of nodes. [5]

All algorithms are difficult to perform on a data structure based on rows and columns (relational databases), as performance would not allow it. Therefore, graphs are always optimal as a data structure when the use case requires highly interconnected data and clusters.

5. Current research

After defining the relevant basic terms and before describing the practical investigation in more detail in section 6, this section provides an overview of the current research regarding data models and the application of graphs to critical success factors.

The following figure shows the current number of papers that deal with the combination of the keywords "start-ups" and "critical success factors" (the papers were selected using the keywords shown in Fig. 3, whether the papers were "Open Access" and were published in the period 2015-2021):

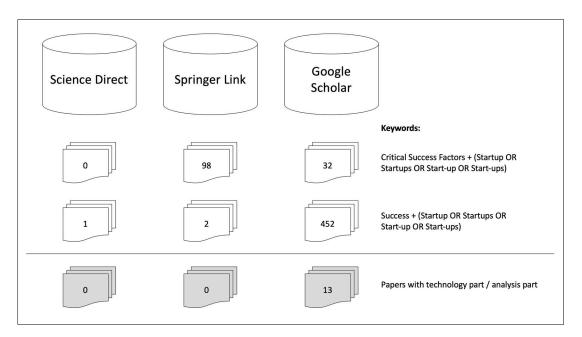


Fig. 3. Systematic selection process of relevant papers for the analysis of current research with regard to analytical methods of start-ups' critical success factors (own creation; as of 03.08.2021)

The figure shows that a total number of 585 papers deal with this topic. Beyond that, however, there are only 13 papers that contain an analysis / technology part that is relevant here. To answer research question 1, the classification in the technological context is crucial. Therefore, the following criteria for examining current research are relevant for this paper:

- 1. Graph-based data models are used.
- 2. Graph databases are used.

3. Contexts are collected or a context-based approach is taken into account according to [1].

The task is to falsify these criteria on the basis of current research. For this purpose, the 13 relevant papers were analysed according to the data source used, the method of analysis and whether the method is context-based:

TABLE 1: ANALYSIS OF CURRENT RESEARCH REGARDING DATA MODELS AND THE APPLICATION OF GRAPHS TO CRITICAL SUCCESS FACTORS DIVIDED INTO DATA SOURCES, ANALYSIS METHODS AND CONTEXT-BASED APPROACHES (YES / NO)

Paper	Data source	Analytical method(s); if one method is given, the main method used is given	Context-based (yes / no)?
[6]	Survey, quantitative survey	Descriptive statistics	No
[7]	Raw data from Crunchbase	K Nearest Neighbour (KNN), Logistic Regression, Bootstrap, Bagging and Boosting, Classification Decision Trees, Classification Decision Trees	No
[8]	Motherbrain	Gated Recurrent Units	No
[9]	Survey	Descriptive statistics	No
[10]	Survey	Descriptive statistics	No
[11]	Survey	Descriptive statistics	No
[12]	Survey	Analytic hierarchy process, Fuzzy Logic	No
[13]	Crunchbase	Logistic regression, Support Vector Machines, Random Forest	No
[14]	Crunchbase	Logistic Regression, Recursive Partitioning Trees (Rpart), Conditional Reference Tree	No
[15]	Crunchbase	Latent Dirichlet Allocation, Neural Network	No
[16]	Crunchbase	Closeness Centrality (Graph), Logistic Regression	No
[17]	Twitter	Latent Dirichlet Allocation, Sentiment Analysis	No
[18]	Coinmarketcap, Websites	Max-Class, Random, Naive Bayes, Random Forest, MLP Logistic, MLP Relu, SVC	No

The following results should therefore be noted:

- 92% of the papers use statistical functions:
 - Descriptive statistics (31%)
 - Logistic Regression (23%)
 - Latent Dirichlet Allocation (15%)
- A graph-like structure was used only once in [15]. Graph databases were not used in any paper.

Another important point in the context of this paper:

• 38% of the papers use Crunchbase as a data source. The relevance of this point becomes clear in the next section.

Based on these findings, it can be stated that in the current research only one paper works with graphs and thus the analysis of critical success factors in the research is not graph-based. Therefore, the two points cannot be completely falsified, but paper [15] can be considered as an outlier in the total number of 585. In addition, the data models are not based on contexts or are carried out context based.

6. Examination

After noting that almost no analysis method of critical success factors in the current research is based on graphs, this section will carry out the experiment based on graphs: The presented method in this section is a) graph-based and b) context-based. In the following, the method is first presented and then the experiment is conducted and analysed.

6.1 Experimental setup

The experimental set-up used for this paper can be outlined as follows:

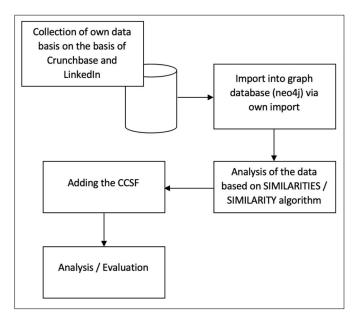


Fig. 4. Experimental setup (own creation)

The figure shows that the data is first collected and then collated. Based on this, they are imported into the graph database via a separate import that was developed for the graph database. This import follows a clear data model (see section 7.2). This is followed by iterating over the data, resulting in the calculation of similarities between the individual data entities (here: start-ups). Afterwards, the CCSF can be added and evaluated via the similarities.

The experimental setup is reproducible and documented step by step.

6.2 Data model

The underlying data model consists of the following components (node (types)):

- "Company": The Company node contains the start-ups' name.
- "Parameter": The parameter node contains descriptive parameters for the respective entity (here: Company).
- "SuccessDefinition": The SuccessDefinition node contains the success definition for the respective start-up (based on parameters).
- "CCSF": The CCSF node contains the context-based critical success factor that applies to the context based on the defined parameters.

The following connections (relations) exist between the individual nodes (types):

- "HAS": A "Company" can have n "Parameters". A "CCSF" can have n "Success-Definitions".
- "IMPACT": A "CCSF" has an impact on n "Companies".
- "SIMILARITY": There are similarities between the "Parameters" and the "Company" (in both directions). The similarity is recorded by the attribute "Value" (0.1; 1 = 100%).

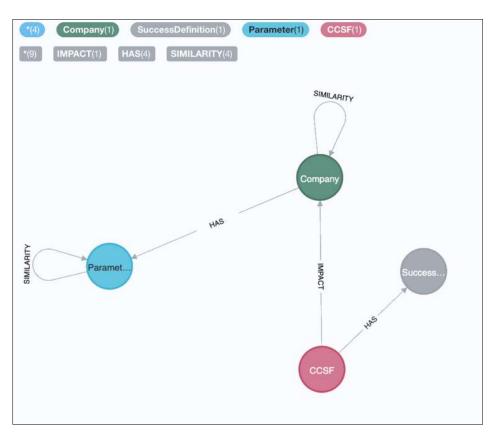


Fig. 5. Data model with "CCSF", "Success-Definition", "Company" and "Parameters" (own creation)

How can this data model be interpreted?

Basically, the data model (schema) is built in such a way that every start-up can be compared with every start-up. In addition, several start-ups can be linked to a context-based critical success factor, which is/was formed dynamically on the basis of the similar parameters and the previously defined success definition. The relation "SIMILARITY" also contains a weighting that allows a relative statement of the similarities and thus creates a dynamic weighting depending on the data situation.

Example of a start-up based on the data model:

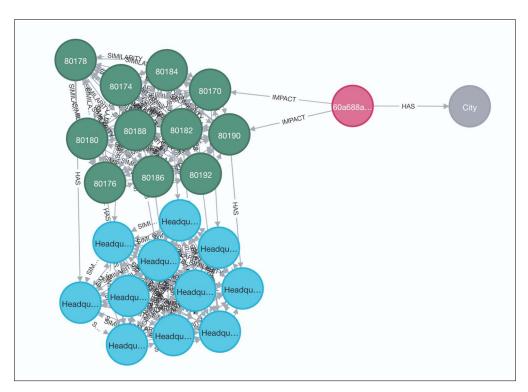


Fig. 6. Exemplary presentation based on an anonymised start-up (own creation)

This exemplary representation shows an anonymised start-up using the data model presented. The similarity values ("Value") are not visible due to the visual representation.

6.3 Data basis

The data basis for this paper was created on the basis of available Crunchbase data (free trial version) and publicly accessible LinkedIn data. A data set of a total of 1,000 German start-ups was created and linked to 1,760 LinkedIn data sets. The data structure used can be seen in the following list. The dataset is normalised and prepared for import (special characters removed). For empty fields, the replacement with "N/A" (not available) was carried out in order to generate a correct import. These fields were omitted in the later analysis then.

The following data fields contain the full package of Crunchbase data that is accessible in the free trial version. The public LinkedIn data fields contain all fields beside the name of the respecting person. No selection of data fields has been made, but all possible data fields have been used.

List of data fields used:

- OrganizationName
- Industry1, Industry2, Industry3, Industry4, Industry5, Industry6, Industry7, Industry8, Industry9, Industry10
- City
- Region
- Country
- Description
- CBRankCompany
- HeadquartersRegions
- DiversitySpotlightUSOnly
- EstimatedRevenueRange
- OperatingStatus
- FoundedDate
- ExitDate
- CompanyType
- Website
- ClosedDate
- Twitter
- Facebook
- LinkedIn
- ContactEmail
- PhoneNumber
- NumberofArticles
- HubTags
- InvestorType
- InvestmentStage
- NumberofPortfolioOrganizations
- NumberofInvestments
- Number of Lead Investments
- NumberofDiversityInvestments
- Number of Exits
- NumberofExitsIPO
- AcceleratorProgramType
- AcceleratorApplicationDeadline
- AcceleratorDurationinweeks
- SchoolType
- SchoolProgram
- NumberofEnrollments
- SchoolMethod
- NumberofFoundersAlumni

- NumberofAlumni
- IndustryGroup1, IndustryGroup2, IndustryGroup3, IndustryGroup4, IndustryGroup5, IndustryGroup6, IndustryGroup7, IndustryGroup8, IndustryGroup9, IndustryGroup10
- NumberofFounders
- Founder1Name, Founder1Linkedin, F1Position1, F1P1Company, F1Position2, F1P2Company, F1Position3, F1P3Company, F1University1, F1University2, F1University3 [LinkedIn data]
- Founder2Name, Founder2Linkedin, F2Position1, F2Company1, F2Position2, F2Company2, F2Position3, F2Company3, F2University1, F2University2, F2University3 [LinkedIn data]
- NameFounder3, Founder3Linkedin, F3Position1, F3Company1, F3Position2, F3Company2, F3Position3, F3Company3, F3University1, F3University2, F3University3 [LinkedIn data]
- NumberofEmployees
- NumberofFundingRounds
- FundingStatus
- LastFundingDate
- LastFundingAmount
- LastFundingType
- LastEquityFundingAmount
- LastEquityFundingType
- TotalEquityFundingAmount
- TotalFundingAmount
- Top5Investors1, Top5Investors2, Top5Investors3, Top5Investors4, Top5Investors5
- NumberofLeadInvestors
- NumberofInvestors
- Number of Acquisitions
- AcquisitionStatus
- TransactionName
- Acquiredby
- AnnouncedDate
- Price
- AcquisitionTerms
- AcquisitionType
- IPOStatus
- IPODate
- DelistedDate
- MoneyRaisedatIPO
- ValuationatIPO
- StockSymbol
- StockExchange
- LastLeadershipHiringDate
- NumberofEvents
- CBRankOrganisation

- CBRankSchool
- TrendScore07Days
- TrendScore30Days
- TrendScore90Days

As additional information to the presented list, it should be mentioned that an overall deduction was made here, which also deliberately includes fields that may not be relevant in the later analysis. However, since the focus of the experiment is on the methodology, these fields were not removed. It should also be mentioned that the note "[LinkedIn data]" indicates that this data is not fetched from Crunchbase, but from LinkedIn. Factually, similar groups (for example, "IndustryGroup", "Top5Investors") were combined in a bullet in the list but were available as individual fields for the import.

6.4 Architecture / System Design

The architecture required for the experimental setup is as follows:

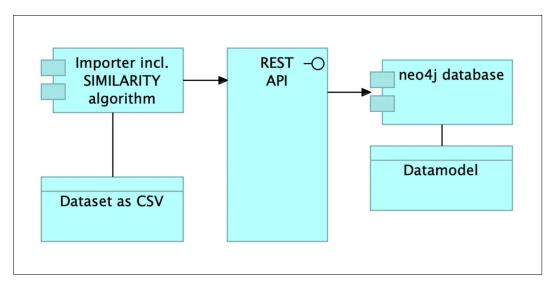


Fig. 7. High-level Architecture / System Design (own creation)

Central components of the architecture are the neo4j database and the importer. In addition, the Representational State Transfer (REST) API and the data set are required. The entire logic, including the algorithm, is contained in the importer. The importer controls the import of the datasets into the neo4j graph database. The import is carried out via the REST API using JavaScript Object Notation (JSON). The JSON contains the corresponding commands (import, calculation) for the database.

6.5 Algorithm / Method

The algorithm and thus the core of the methodology works as follows:

- 1. Read in parameters and company name
- 2. Link parameters to the respective company
- 3. Set similarities between the parameters:
 - a. If the parameter does not occur more than once in the data set (for example, founder positions (the founder can have several professional positions)), the two parameters are compared (two start-ups are always compared directly with each other). A distinction must be made between a numerical and a non-numerical parameter. In the case of a numerical parameter, the comparison is made by the percentage deviation. Since only parameters of the same type are ever compared and the data are normalised, this is permissible. If the parameter is non-numeric, a 100% comparison is performed for short parameters (<= 10 characters). If both parameters are 100% identical, the similarity is 1, otherwise 0. If the string is longer, text similarity algorithms (weighting based on "RAKE" [18] sortByScore() in combination with similar_text() [19] from "PHP") are used to calculate the similarity of the texts and the similarity value is transferred accordingly.
 - b. If the parameter occurs more than once, the same logic as in a) is performed, but only in the respective factually related data set (e.g., universities attended by the founder, top 5 investors, etc.).
- 4. *Set similarities between the start-ups:* The similarity or similarity index between the compared start-ups is set on the basis of the similar parameters. This is calculated via the summed similarities of the parameters between the start-ups and the division by the number of parameters. The highest value is 1 and corresponds to 100%.
- 5. *Set CCSF*: The CCSF is set based on the Success Definition. This can consist of n parameters. These parameters are then searched for among the companies and this subset is compared again. This results in start-ups that are exactly tailored to the CCSF and are compared with each other. Findings might be derived from this subset.

6.6 Implementation

The experiment was carried out on the basis of the algorithm mentioned above and the architecture presented. Various CCSFs were then defined to illustrate the experiment by way of example. The following analyses for examples 1-3 were carried out directly on the database (neo4j, query with Cypher [21]).

Example 1:

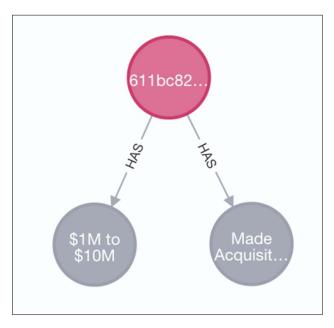


Fig. 8. Exemplary CCSF with definition of success (own creation)

The CCSF (in the following: "CCSF 1") was defined with the following parameters (more precisely: "SuccessDefinition"):

- EstimatedRevenueRange: "\$1M to \$10M"
- AcquisitionStatus: "Made Acquisitions"

These parameters were imported via the importer into the neo4j database as CCSF. These parameters were then decisive for the algorithm presented. After the algorithm has analysed the existing data and identified, compared, and weighted corresponding start-ups, the analysis can take place at database level.

For this purpose, a Cypher statement was used to query which start-ups belonging to CCSF 1 have the most similar parameters and what the weighting between them is (weight = 1 means that the parameters between two start-ups are identical). This can be seen in the following table:

Table 2: Top 10 excerpt on CCSF 1 (the query includes every single parameter of the start-ups belonging to the CCSF)

ParameterType	ParameterValue	ParameterCount	ParameterWeight
HeadquartersRegions	European Union (EU)	101	1
IPOStatus	Private	101	1
OperatingStatus	Active	101	1
CompanyType	For Profit	98	1
Country	Germany	86	1
City	Berlin	59	1
Region	Berlin	59	1
FundingStatus	Early Stage Venture	45	1
EstimatedRevenueRange	\$1M to \$10M	39	1
FundingStatus	Late Stage Venture	26	1

This table shows the first ten most relevant parameters that fit into the context for CCSF 1 and could thus characterise it. "ParameterType" shows the parameter type and "ParameterValue" the corresponding value (see also section 6.3). "ParameterCount" represents the frequency of the parameters that occurred with the weighting via "ParameterWeight" (calculated by the algorithm). It can thus be seen from the table that for CCSF 1 and the associated success definition, the ten most relevant parameters that could lead to success according to the success definition are precisely these.

The following table shows the parameters that are even more specific for CCSF 1 (in this case, "specific" means that parameters that appear after the ten most relevant parameters in the list, but that still occur more often than once in relation to "ParameterCount") - ten parameters in total as well:

Table 3: Top 10 extract according to Parameter Count > 1 for CCSF 1 (the query includes every single parameter of the CCSF of the start-ups belonging to the CCSF)

ParameterType	ParameterValue	ParameterCount	ParameterWeight
Top5Investors2	Accel	2	1
IndustryGroup5	Internet Services	2	1
IndustryGroup2	Professional Services	2	1
Industry2	Internet	2	1
Top5Investors3	Kinnevik AB	2	1
IndustryGroup3	Financial Services	2	1
IndustryGroup2	Lending and Investments	2	1
IndustryGroup3	Transportation	2	1
F1Position2	Partner	2	1
Industry2	Logistics	2	1

CCSF 1 includes a total of 7 start-ups that have been analysed on this basis.

Example 2:

The following example shows the CCSF 2 with the following parameters:

• Number of Employees: "501-1000"

• OperatingStatus: "Active"

A total of 7 start-ups were analysed in this CCSF.

The following two tables (analogous to example 1) can be extracted:

Table 4: Top 10 excerpt on CCSF 2 (the query includes every single parameter of the start-ups belonging to the CCSF)

ParameterType	ParameterValue	ParameterCount	ParameterWeight
HeadquartersRegions	European Union (EU)	101	1
IPOStatus	Private	101	1
OperatingStatus	Active	101	1
CompanyType	For Profit	98	1
Country	Germany	86	1
City	Berlin	59	1
Region	Berlin	59	1
FundingStatus	Early Stage Venture	45	1
FundingStatus	Late Stage Venture	27	1
EstimatedRevenueRange	\$10M to \$50M	26	1

 $\label{thm:ccsf} \mbox{Table 5: Top 10 section by Parameter Count} > 1 \mbox{ for CCSF 2 (the query includes every single parameter of the CCSF) }$

ParameterType	ParameterValue	ParameterCount	ParameterWeight
F1Position1	Co-Founder and CEO	2	1
Industry2	Internet	2	1
IndustryGroup5	Internet Services	2	1
F3University2	Indiana University - Kelley School of Business of MBA ExchangeField Of StudyAccounting & Finance	2	0.26611202481576
F3University2	Indiana University - Kelley School of Business of MBA ExchangeField Of StudyAccounting & Finance	2	0.23354049459313
F1University1	Centre for Digital Technology and Management of Honors Degree Technology ManagementField Of StudyTechnology Management	2	0.2277706193429
F1University1	Centre for Digital Technology and Management of Honours DegreeField Of StudyTechnology Management	2	0.22324964552343

F1University2	Technische Universität München of Master of ScienceField Of StudyMechanical Engineering/Industr ial Management	2	0.21823300320129
F3University2	Indiana University - Kelley School of Business of MBA ExchangeField Of StudyAccounting & Finance	2	0.19403639371381
F1University1	Centre for Digital Technology and Management of Honors Degree Technology ManagementField Of StudyTechnology Management	2	0.18457871554419

Example 3:

The following example shows the CCSF 3 with the following parameters:

- NumberofInvestors >= 3
- TotalEquityFundingAmount >= 1,000,000 (USD)

A total of 60 start-ups were analysed in this CCSF.

 $TABLE\ 6: TOP\ 10\ EXCERPT\ ON\ CCSF\ 3\ (THE\ QUERY\ INCLUDES\ EVERY\ SINGLE\ PARAMETER\ OF\ THE\ START-UPS\ BELONGING\ TO\ THE\ CCSF)$

ParameterType	ParameterValue	ParameterCount	ParameterWeight
HeadquartersRegions	European Union (EU)	99	1
OperatingStatus	Active	99	1
IPOStatus	Private	99	1
CompanyType	For Profit	96	1
Country	Germany	84	1
City	Berlin	57	1
Region	Berlin	57	1
FundingStatus	Early Stage Venture	44	1
EstimatedRevenueRange	\$1M to \$10M	38	1
FundingStatus	Late Stage Venture	26	1

 $\label{thm:conding} \mbox{Table 7: Top 10 extract according to Parameter Count > 1 for CCSF 3 (the query includes every single parameter of the CCSF of the start-ups belonging to the CCSF)}$

ParameterType	ParameterValue	ParameterCount	ParameterWeight
IndustryGroup7	Software	3	1
F2Position2	Co-Founder	2	1
IndustryGroup6	Science and Engineering	2	1
IndustryGroup8	Software	2	1
IndustryGroup3	Data and Analytics	2	1
Top5Investors1	e.ventures	2	1
F1University2	Technische Universität München of Master of ScienceField Of StudyMechanical Engineering/Industri al Management	2	0.34181051016494
F3University2	Indiana University - Kelley School of Business of MBA ExchangeField Of StudyAccounting & Finance	2	0.31714503652284
F3University2	Indiana University - Kelley School of Business of MBA ExchangeField Of StudyAccounting & Finance	2	0.28431905259491
F1University1	Centre for Digital Technology and Management of	2	0.28099534852394

Honors Degree Technology ManagementField Of StudyTechnology Management	

6.7 Analysis / Interpretation

Based on the examples shown, the following conclusions can be drawn:

- There are similarities between the individual parameters in the top 10 parameters due to the amount of data analysed. For example, the parameters "HeadquartersRegions" and "Country" occur frequently, as all analysed start-ups are from Germany and therefore the value "European Union (EU)" and "Germany" are identical for every start-up. The situation is similar with "OperatingStatus" and "IPOStatus: Most start-ups are active and not listed on the stock exchange, too.
- The following parameters, which are not naturally the same due to the data set, differ noticeably in the respective CCSF 1-3. For example, there are differences in "FundingStatus" and "EstimatedRevenueRange" (apart from when "EstimatedRevenueRange" was set as the "SuccessDefinition" in CCSF 1).
- The parameters that are not in the top 10 list of the respective CCSF are particularly interesting. There are clear differences here, which can be seen, for example, in the parameters "IndustryGroup[x]", "F[x]University[y]", "F[x]Position[y]" or "Top5Investors[x]" ("x" and "y" can be replaced by the respective number, as there are several of each of these parameters (see section 6.5: 3b)).
- It can be seen that different start-ups are relevant due to the set definition of success and that different parameters are relevant there, which could have an effect on the respective definition of success (must be checked in further research).

In addition, the experiment allows the following general conclusions to be drawn:

- It has been shown that a CCSF with a fixed definition of success can dynamically filter startups.
- Similar parameters could be found between the start-ups based on their respective contexts.
- Weightings of the various similar parameters could be shown.
- It was also possible to show "fuzzy" similarities, which revealed "soft" similarities between the parameters (parameters ≠ 0,1).

• It can be assumed that different contexts lead to different evaluations, since the parameters that fit the definition of success are different in each case and other start-ups are used (only start-ups that fit the CCSF's definition of success). This thesis needs to be examined in more detail in further research.

6.8 Differences to previous analytical approaches in science

In contrast to current methods / analysis approaches that have been pointed out in section 5, this approach differs in the following features:

- More dynamic calculation of various parameters based on a defined context.
- No limitation of parameters and open for a wide range of parameter types.
- Deliberately no aggregation of data, but retention of the raw data and thus no compression and resulting loss of information.
- Graph-based approach.

7. Conclusion

7.1 Answering the research question

In the following, the research questions posed in section 2 are answered:

- 1. Research question: Are graphs used in the current data models on critical success factors?
 - Answer: Based on the literature review in section 5, it can be stated that only one paper in the current research uses graphs and thus the analysis of critical success factors in the research is not graph-based. Therefore, the two points cannot be completely falsified, but paper [15] can be considered as an outlier in the total number of 585.
- 2. *Research question:* Are contexts collected in the data models or used to examine critical success factors?
 - *Answer: In the* current research, the data models are not based on contexts or are carried out context-based in the sense of [1].
- 3. *Research question:* How does a context-based data model look like using sample data and what insights / success factors can be extracted from it?
 - *Answer:* The presented data model was filled in with sample data and executed by the presented algorithm. The insights gained from the presented experiment are as follows:

- a. There are similarities between the individual parameters in the top 10 parameters due to the amount of data analysed.
- b. It can be seen that different start-ups are relevant due to the set definition of success and that different parameters are relevant there, which could have an effect on the respective definition of success (must be checked in further research).
- c. It has been shown that a CCSF with a fixed definition of success can dynamically filter start-ups.
- d. Similar parameters could be found between the start-ups based on their respective contexts.
- e. It can be assumed that different contexts lead to different evaluations, since the parameters that fit the definition of success are different in each case and other start-ups are used (only start-ups that fit the CCSF's definition of success). This thesis needs to be examined in more detail in further research.

7.2 Critical view

The presented experiment / method offers possibilities regarding the context-based analysis of critical success factors. The logic follows the presented experiment and algorithm. The approach could offer advantages in terms of the relevance of critical success factors, as the selection of CSFs is based on the context and a fixed definition of success. In contrast to current research, not all start-ups are analysed in general, but more fine-grained in terms of their appropriate context, which could lead to better results in terms of the relevant CSFs.

It should be noted that the presented method still has to be proven or tested with regard to its practical suitability (for example for analyses, investment decisions et cetera). This requires further research in this area (see section 7.5).

7.3 Discussion

This section discusses the presented paper. The following points should be noted:

- The approach offers new technical possibilities in the dynamic capture of CSF based on the context (= CCSF).
- This new approach needs to be reviewed and further investigated. In particular, it needs to
 be tested on the basis of different points in time whether the CSFs also become relevant over
 time according to the contextual definition of success (thus in practice).
- Previous approaches are mostly based on aggregations of data (see section 5) and are therefore more generalistic. These approaches (not only at the data-analytical level; see also [1]) thus also offer generalistic CSFs that can apply to all start-ups and are thus, in logic, only ever suitable to a certain percentage (a percentage value is difficult to determine here).

7.4 Limitations

This paper (especially the experiment) is subject to the following limitations:

- 1,000 start-ups from Germany were considered and thus the amount of data was limited.
- The number of parameters was large, but possibly not sufficient. Other parameters besides Crunchbase and LinkedIn (for example, further data of the founders or financial data et cetera) could improve the result even further.
- The data for further definitions of success must be more precise on the basis of further data sources. Here, the limitation was based on the defined parameters.
- The SIMILARITY algorithm presented (first approaches of the algorithm have been developed together with Prof. Dr. Wolfgang Rams in 2016) can still be refined and expanded. Particularly in the text analyses, further progress can still be made in order to carry out further and more precise measurements.

7.5 Outlook

It needs to be shown whether the dynamic CCSF approach offers advantages over a generalistic approach. The experiment in this paper offers indications of this, which need to be verified with historical data comparisons. In addition, further data sources should be connected (for example: turnover data, patent data, et cetera) and the experiment should be conducted with a larger number of start-ups. The larger the number of start-ups per context / CCSF are, the more likely it is to find out which parameters also have an effect on the definition of success in practice.

References

- [1] F. Broszat, M. Fekete, Why context matters for start-ups' critical success factors: the definition of context-based CSF, SIER. 1 (2021) 43–61. https://doi.org/10.52514/sier.v1i1.1.
- [2] E. Ries, The lean startup: how today's entrepreneurs use continuous innovation to create radically successful businesses, 1st ed, Crown Business, New York, 2011.
- [3] A.K. Dey, Understanding and Using Context, Personal and Ubiquitous Computing. 5 (2001) 4–7. https://doi.org/10.1007/s007790170019.
- [4] J.A. Bondy, U.S.R. Murty, Graph theory with applications, North Holland, New York, 1976.
- [5] M. Needham, A.E. Hodler, Graph algorithms: practical examples in Apache Spark and Neo4j, 2019.
- https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=2141755 (accessed August 20, 2021).
- [6] M. Aminova, E. Marchi, The role of innovation on start-up failure vs. its success, International Journal of Business Ethics and (2021). http://ijbeg.com/index.php/1/article/view/60.
- [7] A. Kamal, K. Sabani, Modeling Success Factors for Start-ups in Western Europe through a Statistical Learning Approach, diva-portal.org, 2021. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1561654.
- [8] R.H.A. Stahl, Leveraging Time-Series Signals for Multi-Stage Startup Success Prediction, (2021). https://ethz.ch/content/dam/ethz/special-interest/mtec/chair-of-entrepreneurial-risks-dam/documents/dissertation/master%20thesis/Thesis_Leveraging_Time_Series_Signals_for_Multi_Stage_Startup_Success_Prediction_final.pdf.
- [9] W. Witkittiluck, T. Kortana, The Success Factors of Innovative Marketing Management for SME/Startup Entrepreneurs of Thailand Herbs, Psychology and Education (2021). http://psychologyandeducation.net/pae/index.php/pae/article/view/4974.
- [10] J. Santisteban, J. Inche, ..., Critical success factors throughout the life cycle of information technology start-ups, ... and Sustainability Issues. (2021). https://search.proquest.com/openview/2ceed171b59a9259e0fd354631be8ecf/1?pq-origsite=gscholar\&cbl=4916366.
- [11] J. Santisteban, D. Mauricio, ..., Critical success factors for technology-based startups, International Journal of (2021). https://doi.org/10.1504/IJESB.2021.114266.
- [12] T.V. TRANG, N. VINH, Q. DO, Application of Fuzzy Analytic Hierarchy Process in Prioritizing and Ranking Critical Success Factors of Innovation Startups, (n.d.).
- [13] F. Bento, Predicting start-up success with machine learning, run.unl.pt, 2018. https://run.unl.pt/handle/10362/33785.
- [14] C. Ünal, Searching for a Unicorn: A Machine Learning Approach Towards Startup Success Prediction, edoc.hu-berlin.de, 2019. https://edoc.hu-berlin.de/handle/18452/21141.
- [15] Y.Q. Ang, A. Chia, S. Saghafian, Using Machine Learning to Demystify Startups Funding, Post-Money Valuation, and Success, ... -Money Valuation, and Success, (2020).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3681682.

- [16] M. Bonaventura, V. Ciotti, P. Panzarasa, S. Liverani, ..., Predicting success in the worldwide start-up network, Scientific Reports. (2020). https://www.nature.com/articles/s41598-019-57209-w.
- [17] P.P. Sánchez, J.R. Saura, A. Grilo, Detecting indicators for startup business success: Sentiment analysis using text data mining, idus.us.es, 2019. https://idus.us.es/handle/11441/85637.
- [18] D.D. Dio, DeepSuccess-Predict the Success of Tech Startups, (2019). https://infoscience.epfl.ch/record/263077/files/main deepsuccess is.pdf.
- [19] Donatello-za, RAKE PHP Plus, 2021. https://github.com/Donatello-za/rake-php-plus.
- [20] similar_text, n.d. https://www.php.net/manual/de/function.similar-text.php.
- [21] Cypher, n.d. https://neo4j.com/developer/cypher/.