Factors Contributing to Start-up Valuation:
A Machine Learning Approach

Mariia Garkavenko, Université Grenoble Alpes, LIG, Skopai, France
Tatiana Beliaeva, Skopai, KEDGE Business School, France
Eric Gaussier, Université Grenoble Alpes, LIG, France
Hamid Mirisaee, Skopai, France
Cédric Lagnier, Skopai, France
Agnès Guerraz, Skopai, France

Webinar "Le Business model IP des start-ups"
LES France
May 11, 2021

Background and motivation

Start-up valuation:

FOUNDERS

Is needed to raise money;
Motivates entrepreneurs and sets a value to the efforts and resources they put into a business

INVESTORS

Determines the proportion of shares owned by the investors and the amount of capital each shareholder receives when the company sells

POLICYMAKERS

Allows to allocate funds to those projects that have a potential to be most profitable in the future

Allows to track the effectiveness of strategic decision-making and performance of a start-up

- Because of high risk and often no or little revenues, there is **uncertainty** about the value of start-ups
- Start-up valuations are often determined based on **qualitative** characteristics
- Restricted set of **factors** and inconsistency in their importance ranks
- Human **biases** (Blohm et al., 2020)
- A call for **AI** and data science methods in entrepreneurship research to explore patterns in the big data and predict the events (Lévesque et al., 2020; Schwab & Zhang, 2019)

Research questions

Objective - to identify factors contributing the most to start-up valuation

(1) What factors are significant predictors of start-up valuation?

*(2) What is the importance rank of different **factor groups** to predict the start-up valuation?*

*(3) What **individual factors** from each factor group are most important for start-up valuation?*

Factors determining start-up valuation

Entrepreneur's experience or personality (Macmillan et al.,1985)

Characteristics of the product or service (Knight, 1994)

Strong technology and relationships (Baum & Silverman, 2004)

The quality of team (Franke et al., 2008)

Attractiveness of the industry (Miloud et al., 2012)

External relationships of a new venture

Strategy (Csaszar et al., 2006)

Long-term growth and profitability of the industry (Hall & Hofer, 1993)

Innovative capability (Zheng et al., 2010)

Revenue growth (Block et al., 2019)

Twitter sentiment (Tumasjan et al., 2021)

- **Financial capital**
- **Human capital**
- **Social capital**
- **Industry and market timing**
- **Online legitimacy**
- **Location**

Factors determining start-up valuation

Financial capital	Essential and flexible resource; allows to experiment with new projects and explore new opportunities, protecting from uncertain outcomes (Cooper et al., 1994)
Human capital	Knowledge, skills and experience of founders and management team; an important contributor to new venture performance (Macmillan et al., 1985; Smart, 1999)
Social capital	Relationships of a firm with external partners; valuable for knowledge diffusion and transfer (Florin et al., 2003)
Industry and market timing	Industry size, growth, environmental threats, the level of competition; accessibility to the market and market potential for products (Mason and Stark, 2004)
Online legitimacy	Online visibility and social appreciation; a predictor of new venture survival (Antretter et al., 2018)
Location	Headquarters location, entrepreneurial environment

AI for entrepreneurship research

- Advances in data science and the explosion of the available data
- Processing of large amounts of unstructured and rapidly changing data from many different sources in a fast and unbiased manner
- New research questions and better answers to established questions, addressing of emerging practice needs (George et al., 2016; Tonidandel et al., 2018)
- Exploration of patterns and prediction of events (George et al., 2014; Lévesque et al., 2020; Shmueli, 2010)

AI for entrepreneurship

High-growth firms (Coad & Srhoj, 2019),

New venture survival (Antretter et al., 2019),

Outcomes of crowdfunding start-up pitches (Kaminski & Hopp, 2019)

Personality characteristics of entrepreneurs (Obschonka et al., 2017)

AI for start-up investment and valuation

Investment returns of machine learning algorithms and business angels (Blohm et al., 2020)

CEO emotions and firm valuation in ICOs (Momtaz, 2021)

Prediction of undisclosed start-up valuation (Garkavenko et al., 2021)

Data collection

- Quantitative exploratory approach (Schwab and Zhang, 2019)
- Focus on valuations corresponding to the funding rounds
- Start-up database **Crunchbase** (valuation, funding rounds, team members, industries)

1172 funding rounds with a disclosed valuation for the UK start-ups between 2010 and 2020

- UK business registrar **Companies House** (valuation, people - “officers”)

Additional 1231 funding rounds

- In total, **2403** start-up valuations of 1742 start-ups

Variables

- **409** variables

Variables (features)	Measurement (Crunchbase, Companies House)
Valuation	Pre-money valuation corresponding to funding round in Crunchbase. If not available, obtained from SH01 form of Companies House as $The\ total\ number\ of\ shares\ \times The\ amount\ paid\ on\ each\ share\ - The\ fundraising\ amount$ reported in Crunchbase.

Shares Allotted (including bonus shares)

Date or period during which shares are allotted	From 18/05/2020	To 18/05/2020	
Class of Shares:	SERIES C PREFERENCE	Number allotted	57295
		Nominal value of each share	0.001
Currency:	GBP	Amount paid:	855.093466
		Amount unpaid:	0

Statement of Capital (Totals)

Currency:	GBP	Total number of shares:	217695
		Total aggregate nominal value:	217.695
		Total aggregate amount unpaid:	0

Figure 1. SH01 form containing the information to infer the company valuation (retrieved from Companies House)

Variables

Variables (features)	Measurement (Crunchbase, Companies House, Twitter API, Google Search API)
Financial capital	E.g., funding amount and funding rounds
Human capital	Team size and roles (e.g., number of founders, current and past team members, occupation), experience (e.g., number of past and current appointments, occupation managerial experience), nationality and diversity (e.g., number of foreign officers, female officers, age diversity)
Industry and market timing	E.g., industry, start-up age, number of start-ups founded in the same industry
Online legitimacy	News coverage, social media (e.g., number of tweets, twitter likes, retweets), web visibility (e.g., number of search results)
Social capital	Closeness centrality (two companies are connected if there is a person who worked in both companies)
Location	Region and city of a headquarters

Prediction approach

- ML model – **Gradient boosting** (Friedman, 2001). The method is based on the Classification and Regression Trees and was shown to outperform other methods on related tasks (Caruana & Niculescu-Mizil, 2006). **CatBoost** model, allowing advanced processing of the categorical variables
- **Train-test split**
- **Explainable ML** (Covert et al., 2020a; Mathews, 2019; Molnar, 2019; Roscher et al., 2020). **Feature importance** – predictive power that a feature can provide to the model

Model interpretation

- **Correlations and Univariate Predictors**
- **Features Ablation** (Bengtson and Roth, 2008) – comparing the performance of a model trained on the full set of features and the feature set containing all variables except the studied one
- **Permutation Importance** (Breiman, 2001) – random shuffling of the studied feature's values across dataset and measuring the drop in prediction accuracy on the contaminated dataset compared to the original dataset
- **SHAP** (**SH**apley**A**dditive ex**P**lanations) – shows how much the model's prediction on a particular instance relies on the features' values
- **SAGE** (**S**hapley **A**dditive **G**lobal importanc**E**) (Covert et al., 2020b) – estimates the usefulness of a feature for the model's accuracy on the whole dataset

Results: Feature group contributions

Table 1. Feature Groups Impact Analysis

Rank	Feature Group	Group Model (R^2)	Feature Group Ablation (ΔR^2)	Group Permutation Importance (ΔR^2)	SAGE
1	Financial capital	0.452	0.062	0.331	0.134
2	Industry and market timing	0.268	0.028	0.104	0.057
3	Human capital	0.379	0.018	0.086	0.055
4	Online legitimacy	0.312	0.035	0.088	0.046
5	Social capital	0.032	0.001	0.000	0.001
6	Location	-0.012	0.003	0.006	0.002
	All Features	*0.578*			

Notes. SAGE - Shapley Additive Global importancE. The feature groups are ranked by their level of importance for start-up valuation according to the results obtained with SAGE.

- The average performance of the model $R^2 = 0.578$
- According to all methods, **financial capital** is the most critical group of factors
- **Human capital** and **industry and market timing** have almost the same power, and **online legitimacy** is slightly less valuable (SAGE)
- **Social capital** and **location** groups are of low importance in all methods

Results: Individual features contributions

Rank	Correlation Analysis		SAGE approach	
	Variable	Spearman ρ	Variable	SAGE
		Financial Capital		
1	Max. funding amount	0.778	Max. funding amount	47.88
2	Mean funding amount	0.769	Mean funding amount	37.63
3	Total funding amount	0.762	Last funding amount	30.85
4	Last funding amount	0.753	Total funding amount	30.38
5	Funding rounds	0.346	Funding rounds	0.510
		Human Capital		
1	SD time in company	0.486	SD time in company	7.35
2	Assigned officers	0.481	Mean officer age	5.71
3	Active officers	0.448	Max officer age	5.06
4	Max. total appointments	0.388	Featured team	4.40
5	Max. past appointments	0.386	Active officers	3.95
		Industry and Market Timing		
1	Start-up age	0.529	Start-up age	26.47
2	Mean industry costliness	0.367	Mean industry costliness	8.21
3	Max industry costliness	0.354	Young startup	6.43
4	Min industry costliness	0.294	Max industry costliness	5.01
5	Max funding raised in industry	0.217	Min industry costliness	4.78
		Online Legitimacy		
1	#News	0.363	Max Twitter likes (lifetime)	4.26
2	Mean Twitter likes (6m)	0.283	Max. Twitter likes (6m)	3.28
3	Search results from own domain	0.271	#News	3.13
4	SD Twitter likes (6m)	0.266	Mean Twitter likes (6m)	2.31
5	Max Twitter likes (lifetime)	0.264	SD Twitter likes (6m)	2.10

Table 2. Individual Features Impact Analysis from Each Feature Group

Notes: All features in the table are significant at p<0.001. SAGE Shapley Additive Global importancE. For readability, all SAGE values are multiplied by 1000.

Case study

Start-up "Fuse Universal" on the date 2018-05-01

The plot shows how features push the model's prediction from the base value – the average model prediction on the training dataset

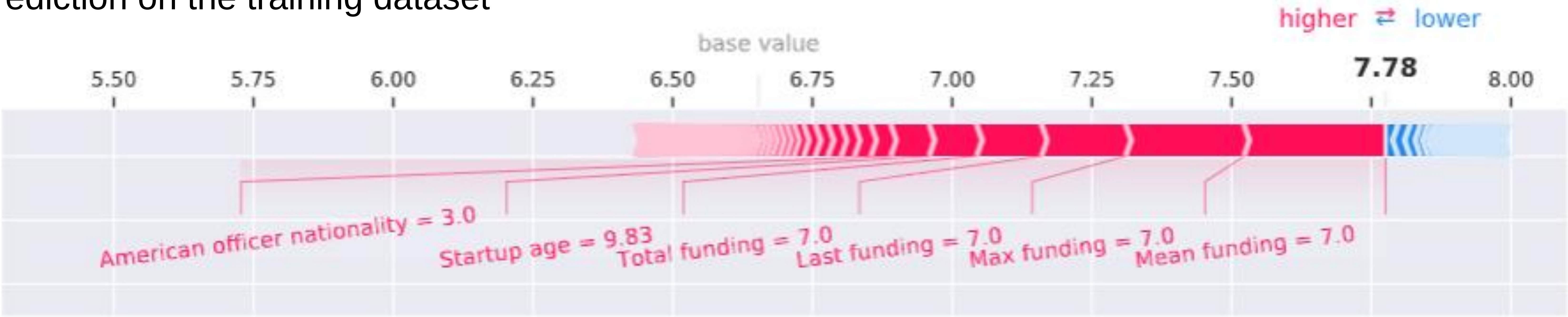

Figure 2. SHAP force plots illustrating features that push model's prediction from base value

- ✓ 1 funding round of $10M (mean / max / last / total)
- ✓ The fact that start-up is mature
- ✓ Has several Americans in the team

Push the model's prediction higher : $10^{7.78}$ = $60M

According to Companies House, in reality, the start-up was valued $ 55M at that moment

Contribution

- Proposition of a ML approach to the valuation problem of start-ups, based on quantitative data and hundreds of different factors
- Comparative insights on the contributions of factor groups to predicting the valuation of start-ups
- Going beyond analyzing feature groups, and empirically showing which individual features are most important for predicting the valuation
- Practical illustration of how factors can be used to explain valuation prediction by pushing it up or down from the average value

Thank you for your attention !