

# Faktory ovlivňující úspěch startupů: Logistická regresní analýza<sup>1</sup>

## Teoretický úvod

**Startup** je definován jako dočasný projekt hledající opakovatelný byznys model generující výnos pro investory. Úspěch startupu může být formou:

- **IPO** (Initial Public Offering): První veřejná nabídka akcií, kdy společnost vstupuje na burzu. Je považována za vrcholný úspěch startupu.
- **Akvizice** (M&A): Odkoupení startupu jinou, obvykle větší společností. Pro investory i zakladatele jde o formu zhodnocení investice.

Neúspěch je pak buď likvidace společnosti nebo neschopnost generovat výnosy pro investory.

Odborná literatura (Gompers et al., 2010) naznačuje, že úspěch není náhodný, ale koreluje s lidským kapitálem zakladatelů. Zkušenost zakladatelů hraje roli. Studie Giot a Schwienbacher (2007) pomocí analýzy přežití (model konkurenčních rizik) ukazují, že dynamika financování (počet kol) zvyšuje pravděpodobnost úspěšného exitu, avšak příliš mnoho kol bez lepšího výsledku zase silně koreluje s likvidací projektu.

Cílem práce je identifikovat prediktory, které souvisí s úspěchem startupů.

## Metodologie a soubor

Pro analýzu byl využit dataset z platformy Kaggle, obsahující data o 100 000 startupech. Data byla pravděpodobně vytvořena za účelem LLM (strojového učení) a datové analýzy.

Závislá proměnná je úspěch (success), 1 = startup byl úspěšný; 0 = neúspěch. Tu jsem vytvořil transformovaním proměnné Outcome (kategorická proměnná o 3 kategoriích – IPO, akvizice a neúspěch)

Nezávislé proměnné jsou:

- Kardinální: **velikost týmu, roky zkušeností zakladatele, průměrné měsíční výdaje, velikost cílového trhu, počet vln financování** pomocí sbírek (fundraising)
- Nominální: **pozadí zakladatele** (z akademického prostředí, první startup, několikátý startup, z Big Tech<sup>2</sup>; **Sektor** startupu (A.I., Klima, Krypto, Digitální trh, Zdravotnictví, Finanční online služby, Softwarové služby)

Kvůli povaze závislé proměnné (dichotomická) byla zvolena pro analýzu **Binominální logistická regrese**. Na rozdíl od klasické lineární regrese nevyjadřuje vztah přímkou, ale pomocí tzv. logitové funkce, která transformuje lineární kombinaci faktorů na pravděpodobnost v rozmezí 0 až 1. Výpočet je formou poměrů šancí.

---

<sup>1</sup> Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=804>.

<sup>2</sup> Zakladatel, který pochází z velké firmy nebo korporátu.

K výpočtu byl použit statistický software R.<sup>3</sup>

## Analýza a výsledky

Byla testována absence **multikolinearity** (pomocí **VIF**), **linearita vztahu** mezi spojitými prediktory (**Box-Tidwell**) a logitem závislé proměnné a **Cookova vzdálenost odlehlých hodnot**. Všechny předpoklady pro Bin. Log. Regresi byly splněny.

Model logistické regrese identifikoval několik statisticky významných prediktorů úspěchu startupů. Statistická významnost byla testována **Waldovým testem**.

Tabulka. č.1: Výsledky logistické regresní analýzy.

	Odds Ratio	p-hodnota
(Intercept)	0.117	< 0,001
Počet vln sbírek (funding_rounds)	1,449 (+ 44,9 %)	< 0,001
Roky zkušeností zakladatele (founder_experience_years)	1,062 (+ 6,2 %)	< 0,001
Velikost týmu (team_size)	1,002 (+ 0,2 %)	< 0,001
První startup zakladatele (backgroundfirst_time)	1,038 (+ 3,8 %)	0,039
Několikátý startup zakladatele (serial_founder)	1,037 (+ 3,7 %)	0,111
Zakladatel z Big Techu (ex_bigtech)	1,014 (+ 1,4 %)	0,47
Měsíční náklady (burn_rate_million)	1	0,113
Velikost cílového trhu (market_size_billion)	1	0,548
Sektor (Crypto, Health, SaaS, Fintech, Ecommerce, ref.: A.I.) <sup>4</sup>	0,968 - 1,036	0,15 - 0,54

Jako nejvlivnější prediktor se ukázal **počet vln sbírek**. Jedno další kolo sbírky na startup zvyšuje šanci na jeho úspěch o necelých **45 %**. Také roky zkušeností zakladatele zvyšují šanci na úspěch startupu o 6,2 % o jeden rok navíc. Jako méně silný prediktor je také velikost týmu, kde za 1 osobu navíc se šance na úspěch startupu zvyšuje o 0,2 %. Všechny tyto prediktory jsou statisticky významné.

Naopak prediktory jako velikost cílového trhu nebo sektor, ve kterém startup působí, nejsou statisticky významné, a tudíž nemají vliv na úspěšnost startupu.

Pod hranicí  $p < 0,05$  je také proměnná první startup zakladatele, která ukazuje, že zakladatelé svého prvního startupu mají vyšší šanci, že jejich startup uspěje než jejich referenční skupina, což jsou zakladatelé z akademického prostředí.

Při hodnocení modelu nám vyšlo, že náš model dosahuje hodnoty **0.0795 (McFadden Pseudo R<sup>2</sup>)** a to znamená, že se o danou hodnotu relativně zlepšil než nulový model s žádnými prediktory.

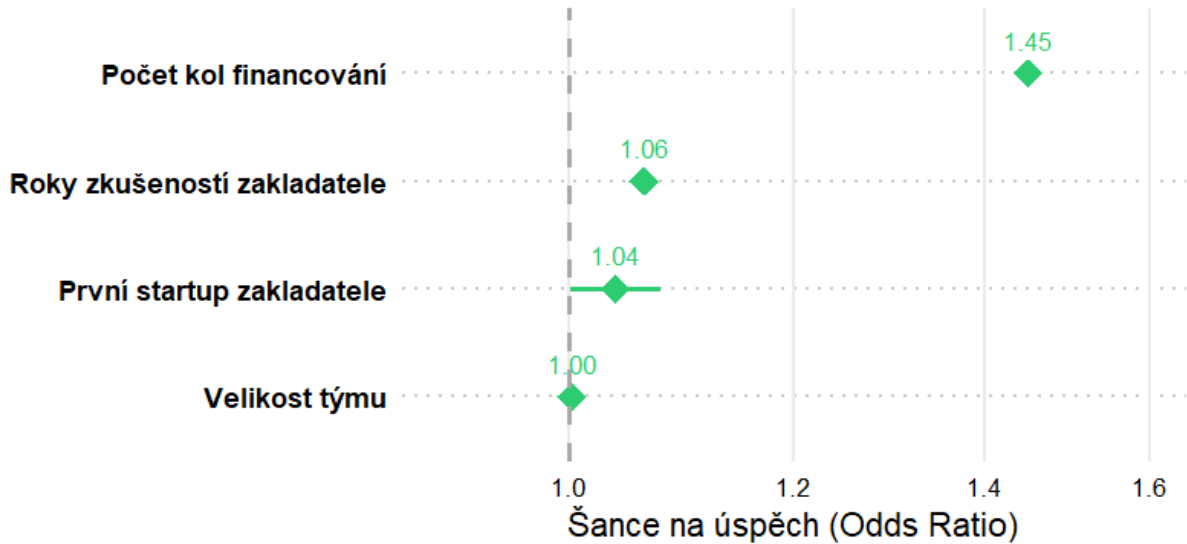
<sup>3</sup> K analýze byly použity R knihovny: dplyr; car; broom; pROC a ResourceSelection.

<sup>4</sup> Jako referenční skupina byl zvolen sektor umělé inteligence.

**Hosmer-Lemeshow** test vyšel **0.1411**, nemůžeme tudíž vyvrátit nulovou hypotézu, což znamená, že model dobře sedí na naše data.

Graf č. 1: Forest plot se statisticky významnými prediktory.

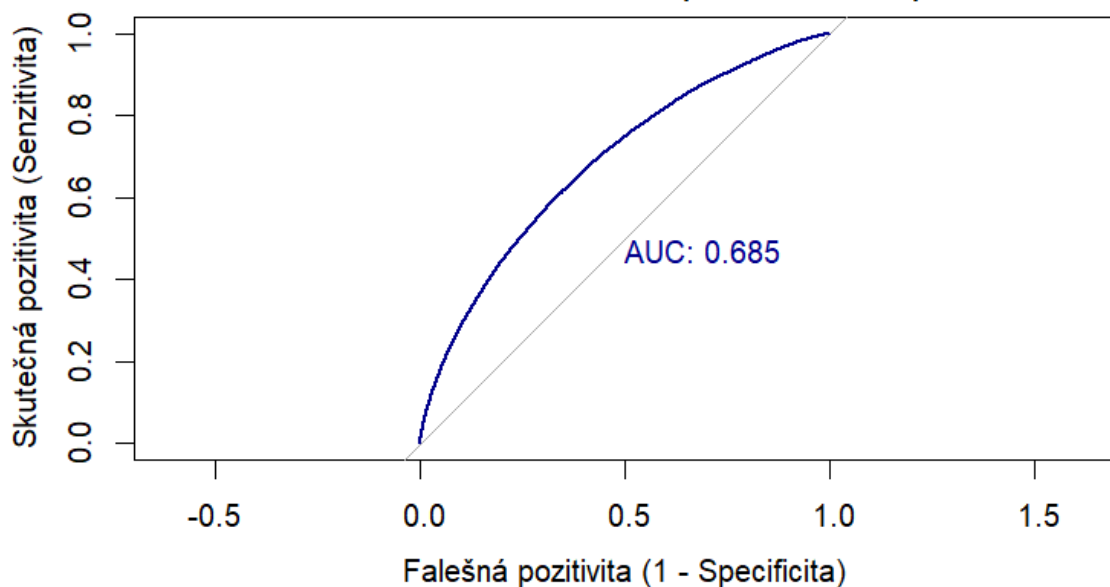
### Forest plot s 95% CI



**AUC** nám vyšlo **0,6851**, což znamená, že když bychom vybrali jeden případ (startup), model má 68,51% šanci, že ho zařadí správně jako úspěšný nebo neúspěšný. ROC křivka je poměrně plochá a nemá zaklenutý levý roh. Celkově to znamená, že model predikuje lépe než hod mincí, ale má problém dosáhnout vyšší senzitivity (označit správně úspěšné startupy) a zároveň nezvýšit příliš specifitu (falešně pozitivní).

Graf č. 2: ROC křivka.

### ROC Křivka modelu (AUC = 0.685 )



## Závěr

Logistická regrese nám umožnila izolovat vliv jednotlivých faktorů při očištění o ostatní proměnné. Díky testu Hosmer-Lemeshow víme, že model je dobře zkalibrován. Přestože model nedokáže s vysokou přesností předpovídat úspěch jednotlivých projektů, identifikoval klíčové faktory, které statisticky významně zvyšují šance na úspěch.

Kdyby šlo o opravdové data, znamenalo by to, že v podstatě můžeme potvrdit tvrzení o roli lidského kapitálu jako zkušenosti v daném odvětví. Rovněž bychom mohli potvrdit, že počet sbírek financování na projekt významně souvisí s úspěchem startupů.

U prediktoru počtu kol financování může jít o korelaci, ne kauzalitu. Vyšší počet sbírek může značit, že startup byl jednoduše kvalitní už od začátku a úspěšné financování tuto kvalitu spíše jen potvrzují, než že by zapříčinily úspěch celého startupu.

*Odkazy na literaturu:*

Gompers, P. A., Kovner, A., Lerner, J., & Scharfstein, D. S. (2010). Performance persistence in entrepreneurship. *Journal of Financial Economics*, 96(1), 18-32.

Giot, P., & Schwienbacher, A. (2007). IPOs, trade sales and liquidations: Modeling venture capital exits using survival analysis. *Journal of Banking & Finance*, 31(3), 679-702.

*Dataset:*

<https://www.kaggle.com/datasets/dhrubangtalukdar/startup-funding-and-outcome-dataset/data>