The American Journal of Management and Economics Innovations
ISSN 2693-0811 | Open Access

TYPE Original Research
PAGE NO. 46-53
DOI 10.37547/tajmei/Volume07Issue08-05

Check for updates

Algorithmic Identification of Relevant Investors Using Machine Learning

Anna Mastykina
Founder, Taskinfinity.com, FundWise LLC Buenos Aires, Argentina

OPEN ACCESS
SUBMITTED 22 July 2025
ACCEPTED 27 July 2025
PUBLISHED 12 August 2025
VOLUME Vol.07 Issue 08 2025

CITATION

Anna Mastykina. (2025). Algorithmic Identification of Relevant Investors Using Machine Learning. The American Journal of Management and Economics Innovations, 7(8), 46–53. https://doi.org/10.37547/tajmei/Volume07Issue08-05

Abstract: In this article, the problem of the low efficiency of traditional cold communications with venture capital funds is examined. The relevance of the study is determined by the need to develop automated tools for targeted search of relevant investors capable of overcoming the limitations of warm recommendations and expanding access to capital for startup teams without an extensive network. The aim of the paper is to demonstrate an algorithmic approach based on machine learning methods for identifying relevant investors and to investigate the integration of ML ranking with a disciplined multistep-outreach strategy. The novelty lies in the use of a multilayer feature architecture combining an investment graph, thematic embeddings, soft signals from public channels, and dynamic indicators of fund activity, as well as in the construction of a controlled cycle of cold communications with two follow-ups in each three-day window. The obtained results confirm an increase in the efficiency of the cold channel: algorithmic selection enabled maintaining an open rate at the level of 74–80%, a reply rate in the range of 10–17%, and provided 96 scheduled calls per quarter without a single warm recommendation. The integration of the ML ranking model with a structured cadence strategy increases the controllability of the process, turning fundraising from a lottery into a repeatable business process with continuous model learning on feedback data. Practical implementation includes not only the development of an investor ranking model but also the creation of infrastructure for large-scale mailings: configuration of mail domains, optimization of message templates, A/B-testing, and integration with meeting-scheduling tools. This allows startups to systematically increase open, click, and reply rates as well as conversion into negotiations. The article will be useful to startup founders, venture analysts, and

fundraising specialists seeking to improve the efficiency of cold communications with investors.

Keywords: algorithmic investor identification, machine learning, cold outreach, fundraising, investor ranking, multistep outreach, email marketing, link prediction

Introduction

Mailboxes of venture fund partners are overflowing: the average open rate of business email campaigns fluctuates between 17 and 28%, depending on the industry, while click-through rates in most sectors barely reach 2% (Campaign Monitor, n.d.). The low conversion is exacerbated by the very method of investor selection. This distribution turns traditional manual research into a vicious circle: to get on the radar of funds, a warm intro is needed, and intros are obtained only when the company is already visible on the market or the founder has previously sold a successful business.

The present article pursues two objectives. First, it demonstrates that algorithmic, machine-learning-based investor search can radically increase the efficiency of the cold channel even for teams without an elite network. Second, it shows how combining ML ranking with a disciplined multistep outreach turns fundraising from a lottery into a controllable process: in the author's campaigns, this has already yielded an open rate of 78%, a response rate of 13%, and 96 fund calls in three months without a single warm intro. Thus, the key thesis of the work is formulated: in the era of available data and cloud models, the success of capital raising is determined not so much by the circle of personal acquaintances as by the quality of the algorithm that places the right investors at the top of your funnel.

Materials and Methodology

The study of algorithmic identification of relevant investors is based on the analysis of eleven sources, including industry reports on email marketing, Campaign Monitor, and SendGrid, cold-email campaign statistics, Chatelaine, Pipeful, and Calendly, and the author's campaign data. The theoretical basis comprised works on the construction of investment graphs and link prediction using Crunchbase (Piloterr), analysis of the role of soft signals from public channels (Kaiser & Kuckertz), and dynamic metrics of fund activity (Venture Capital Association; Te et al.).

Methodologically the study combined comparative analysis of deliverability and conversion—comparing average open and reply rates with industry benchmarks to results of algorithmic selection; extraction of multilayer features (Piloterr), thematic embeddings from startup descriptions, clustering of investor tweets and accounting for dynamics of rounds and headcount (LinkedIn; Venture Capital Association; Te et al.); systematic review of cold-email campaign practices—optimization of templates following the fifteen-words plus one metric plus links to Deck and Calendly principle and configuration of DMARC/SPF/DKIM DNS to improve domain reputation (McGee).

Results and Discussion

Early-stage fundraising operates almost under the same laws as the classic B2B sales funnel. At the entry stage, a bulk of investment partner contacts is followed by email opening, reply, scheduling a call, and, ideally, a term sheet. The problem is that statistics for cold emails remain ruthless: on average, the median reply rate drops to 8.5–10% for most campaigns; only the top quartile of sequences exceeds the 20% reply threshold (Chatelaine, 2024). With such conversion, a founder relying on random responses is doomed to receive few calls after hundreds of emails, while each day of delay reduces the traction effect and intensifies competition for funding attention.

For a long time, this statistic was compensated by personal intros; however, the math here is also relentless: some deals come from former colleagues or business acquaintances, others from secondary recommendations. For teams without an elite network, this scenario creates a systemic gap between the need for capital and the real ability to reach a partner.

The algorithmic approach fundamentally changes the mechanics of the top of the funnel. In the author's campaigns, the open rate is consistently maintained at 74–80%, and the reply share fluctuates between 10% and 17%, as shown in Fig. 1. Each email also converts into a call, yielding a total of 96 meetings over one quarter without a single warm intro. The main effect manifests not only in the growth of individual metrics but in the controllability of the process: instead of sending emails into the void, a clear sequence of touches is formed where each event—opening, click, reply—returns to the model as a training signal.

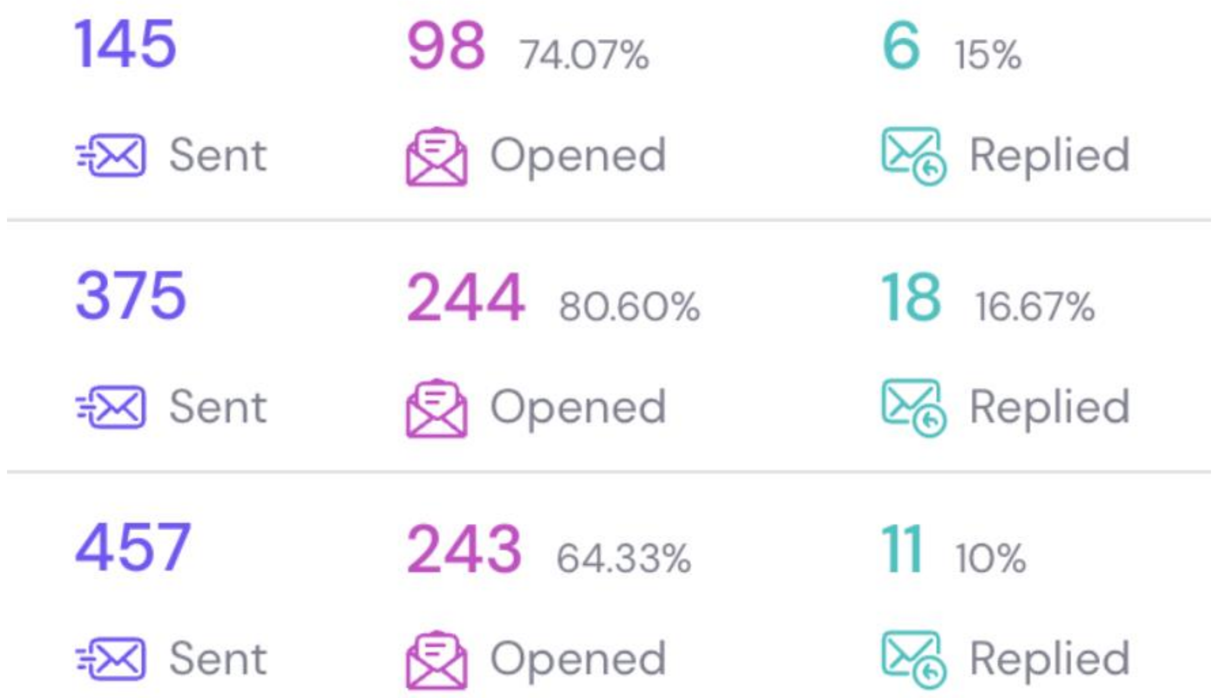

Fig. 1. Results of one of the author's outreach campaigns (compiled by the author)

The key to such robustness is data and discipline. First, sample breadth: for the algorithm to select a relevant top ten, it requires a pool of several hundred funds enriched with features of stage, ticket size, syndicates, and public activity. Second, update frequency: the investment graph changes weekly, and the model drifts if new deals are not incorporated. Finally, strict adherence to cadence — two follow-up touches every three days — retains the investor in the funnel longer than a single email and provides the algorithm with additional feedback points. Without this operational discipline, even the most accurate scoring collapses into statistical noise; with it, fundraising transforms from a one-off campaign into a repeatable process where iterative model improvement directly reflects in top and mid-funnel metrics.

Algorithmic investor identification relies on a four-layer data corpus that combines both structured deal registries and soft signals from public sources. The foundation is an investment graph: historical rounds, exits, and co-investors are collapsed into a directed multigraph where nodes represent funds and startups, and edges are labeled by deal type and date. The public Crunchbase export alone contains approximately 3.5 million company records, each with round → investor and startup → exit linkages, yielding millions of edges for training link prediction and ranking models (Piloterr, n.d.). This is supplemented by commercial datasets: for example, QuantumLight builds scoring based on 700,000 VC-backed companies, demonstrating that graph scale is critical for recommendation quality (Thornhill, 2025).

The next layer comprises portfolio descriptors. For each company, the model extracts thematic and technological features from descriptions, tags, and stack words; texts are converted into embeddings and aggregated into a fund's niche temperature.

To capture implicit shifts in capital focus, the graph is enriched with partners' public theses. Features include keywords and emotional markers from the Twitter stream: a recent sample of 994,969 tweets from 822 investors enabled the authors to identify clusters such as infrastructure AI or climate-tech that directly correlate with subsequent checks (Kaiser & Kuckertz, 2024).

Finally, activity signals impart temporal sensitivity to the model. The average fund seeks to close only seven deals per year — a low throughput through the funnel — and the probability of a response increases sharply if scoring marks the partner as having recently closed a round (Venture Capital Association, 2023). Hiring pace is added to the dynamic features: combined Crunchbase-LinkedIn analyses show that headcount growth and Series A probability are statistically linked, thus a monthly sampling of team profiles enters the dataset as a proxy traction metric (Te et al., 2023).

Such a multilayer corpus—deal graph, thematic embeddings, public theses, and live activity indicators — provides the model with sufficient information to rank investors by current relevance with high stability rather than by past-experience stereotypes.

As noted in the Introduction, the average business email is opened by only 17–28% of recipients and clicked by approximately 4.48% (Send Grid, n.d.). Against this background, every structural improvement of the email becomes critical: in the author's internal campaign sample, the template redesign increased the open rate while maintaining the reply rate. The key element of this dynamic is the first text block, limited to fifteen words: it simultaneously states the essence of the product and demonstrates a concrete result. Due to its high information density, the recipient decides to open the attachment before reading the rest of the email.

This is followed by a single bullet highlighting the key growth metric: conversions drop sharply if the list is expanded beyond one figure because attention shifts from the call to action to details. Brevity is offset by a hyperlink to a ten-page Pitch Deck: email analysis showed that the metric + document combination nearly doubles the probability of entering the data room compared to the metric alone, since the investor receives both the incentive and a convenient format for hypothesis verification. The first interaction finishes with an explicit connection to Calendly, giving the choice to pick a slot in the present week; a Focus Digital study says that having such a tool increases total deal conversion to 0.21% (McGee, 2024). All three — the brief intro, one strong figure, and two clickable items — the Deck and the calendar — make up a simple but enough info package that changes the balance.

B2B campaign experience shows that a single email seldom closes the investor: Pipeful analytics on a sample of 11 million cold emails records a reply-rate increase of almost 50% after the first reminder, meaning that it is the follow-up that transforms a formal contact into a dialogue (What Are B2B Cold Email Response Rates?, 2024). Longer sequences amplify the effect: a Calendly study based on 300,000 emails demonstrated that a three-touch series raises the cumulative reply rate from 1% to 9%, and with seven touches the metric climbs to 27% (Batrawy & Cottle, 2024). At the same time, nearly half (48%) of salespeople do not make repeat calls; however, 93% of converted leads are often achieved only after the sixth cold-call attempt, as shown in Fig. 2.

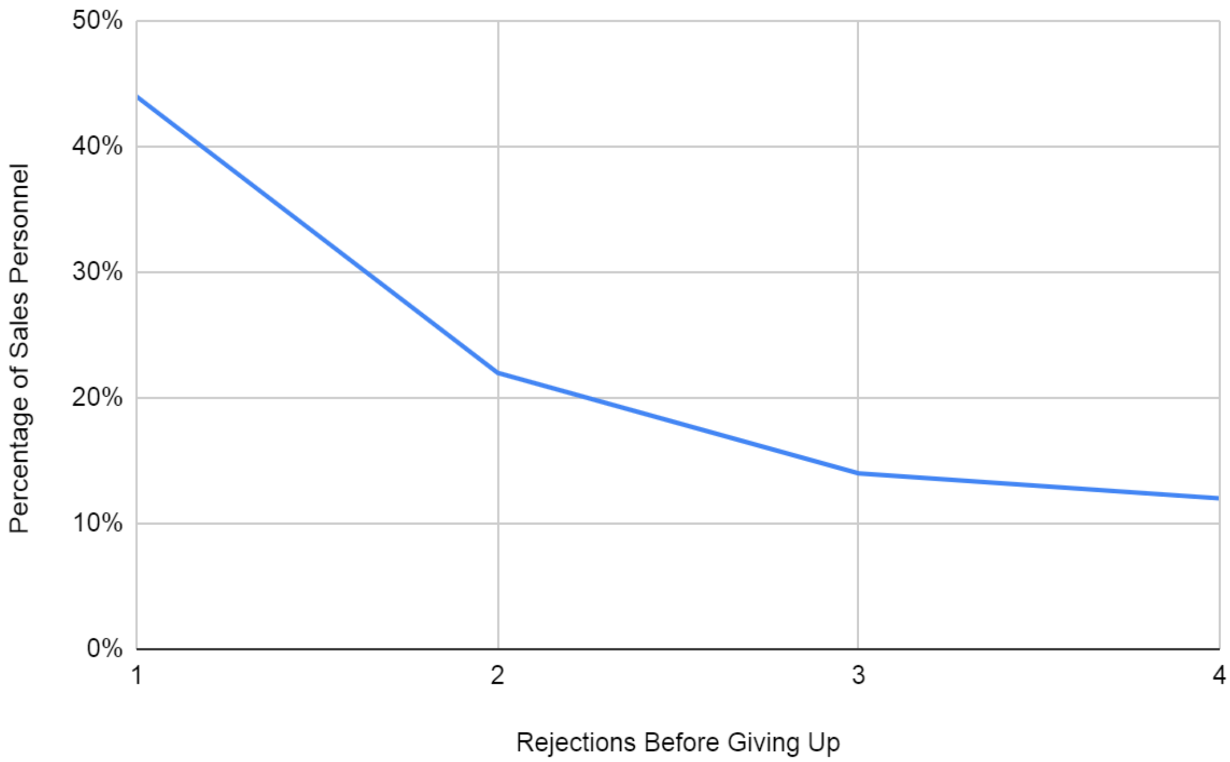

Fig. 2. Percentage of sales personnel who give up after several rejections (Batrawy & Cottle, 2024)

Let us consider an example of an algorithm implemented with TaskInfinity. First, the team formulates a brief description of the startup — one or two sentences about the sector, stage, and key product features, so that the system can correctly understand the target filters for investor search. Then, on the investor-search platform, mandatory filters are set for industry (Industry), round stage (Stage), and market (Market), with additional optional filters such as geography, investor type, and diversity criteria.

Next, an initial shortlist of investors is formed, and each participant is assigned a weight based on four metrics: the number of deals in selected industries and related to LLM (weight 0.1), the number of deals at selected stages (0.4), the number of deals at selected markets (0.4), and the number of deals by diversity (0.1). If any metrics are missing, the system retains proportions and reallocates 100% of the weight among the remaining filters. After calculating the weights, investors are categorized into three groups: the "green" category for a total weight of 45% and above, the "yellow" category for 15–44%, and the "red" category for 1–14%.

Within each category, a second prioritization is carried out based on three indicators with equal coefficients (0.33 each): the ratio of recent investments to total investments, leading participation in selected stages, and the share of total investments from filters. By summing the percentages from the first and second prioritization, the system generates the final ranking of investors within the "green," "yellow," and "red" groups in descending order of the total score.

After this, investors with high and medium relevance are selected from the list. The database is enriched with contact details through the "Get Investor Contacts" interface, which pulls email, LinkedIn profile links, and other required fields. Investors in the "green" and "yellow" categories are marked, and a ready CSV file is exported with columns such as "First Name," "Last Name," "Company," "Email," "LinkedIn," "Website," and other details for subsequent mass mailing.

At the same time, the infrastructure for cold email campaigns is set up: additional domains are registered, mailboxes are created via Google Workspace, 301 redirects and DNS records (SPF, DMARC, DKIM, CNAME tracking) are configured. Over one to two weeks, a "warm-up" is conducted by sending small test emails to improve the reputation of the new domains. LinkedIn accounts are prepared separately: profiles are completed, the contact network is grown, and basic activity is maintained to avoid blocks. During cold outreach, limits are followed — no more than 20 invitations per day and 100 messages per week.

Finally, sequential messaging scripts are created. In LinkedIn, an invitation to connect is first sent mentioning mutual contacts and requesting advice, followed by a short proposal for a call after acceptance; two days later, if no reply is received, a follow-up with an expanded bio is sent; another two days later, a polite apology message is sent asking for expert feedback. In email campaigns, the first message contains an introduction to the team, key metrics (ARR, YoY, GM, LTV/CAC), information about past investments, the round status, and a link to schedule a call; after three days, the first follow-up is sent with “Hope you are having a great week”; another three days later, the second follow-up is sent with a brief repetition of metrics and a link to the Pitch Deck. Once templates and contact lists are prepared, they are uploaded to the chosen mailing tool, and automatic sequences are activated with set limits on outbound messages and follow-up ratios. During the campaign, key metrics — open rate, response rate, and scheduled meetings — are monitored, and adjustments are made to messaging, segmentation, or sending frequency. A/B tests on subject lines and offers are conducted to optimize performance. After collecting the data, strategy and scripts are adjusted, and outreach continues with the updated database. This systematic approach allows for continuously finding the right investors, warming up domains and accounts, sending personalized messages, and quickly optimizing tactics based on actual results.

It was this algorithm that enabled the author to run a few successful campaigns - data from one of them is shown in Fig. 3.

	May 19 - May 25	May 26 - June 1	June 2 - June 8
1-st Email			
Leads outreached	936	1331	900
Emails opened	465	841	584
%	49.68%	63.19%	64.89%
Emails clicked	153	367	283
%	32.90%	43.64%	48.46%
Emails replied	34	49	40
%	7.31%	5.83%	6.85%
2-nd Email (FU in 3 days)			
Emails sent	687	604	793
Emails opened	169	253	400
%	24.60%	41.89%	50.44%
Emails clicked	25	137	171
%	14.79%	54.15%	42.75%
Emails replied	3	12	17
%	1.78%	4.74%	4.25%
3-rd Email (FU in 3 days)			
Emails sent	59	653	818
Emails opened	25	335	509
%	42.37%	51.30%	62.22%
Emails clicked	0	79	231
%	0.00%	23.58%	45.38%
Emails replied	1	13	20
%	4.00%	3.88%	3.93%
4-th Email (FU in 3 days)			
Emails sent	59	502	496
Emails opened	25	115	243
%	42.37%	22.91%	48.99%
Emails clicked	0	0	0
%	0.00%	0.00%	0.00%
Emails replied	1	5	7
%	4.00%	4.35%	2.88%
	14	17	21

Fig. 3. Outcomes of one of the author's campaigns (compiled by the author)

The author of the article first formulated a brief description of the startup and configured filters by industry, stage, geography, and investor type, after which a list of relevant contacts was obtained and new domains were warmed up by test mailings. In the period from 19 to 25 May, a total of 936 emails were sent, of which 465 were opened (49,7%), 153 were clicked (32,9%), and 34 generated a reply (7,3%). A follow-up three days later in the same week, on 687 sent messages, yielded 169 opens (24,6%), 25 clicks (14,8%), and 3 replies (1,8%). Two subsequent reminders in small batches of 59 messages each confirmed the validity of the scenario by adding one or two additional responses.

After analysing the initial results, the author adjusted the subject line and body text of the first email, enhanced personalization, and adapted sending times. In the second week, of 1331 emails sent, 63,2% were opened, 43,6% were clicked, and 5,8% generated replies. In the first follow-up of 604 sent messages, there were 253 opens, 137 clicks, and 12 replies. The third wave of 653 emails produced 335 opens, 79 clicks, and 13 replies, while the fourth reminder resulted in 115 opens and 5 replies. This yielded 17 scheduled calls, exceeding the first week's results by one third.

In the third week, the author continued fine-tuning follow-ups and segmenting by time zone. Of 900 emails sent, 584 were opened, 283 clicked, and 40 replied. The second message achieved 400 opens, 171 clicks, and 17 replies. The third wave of 818 emails recorded 509 opens, 231 clicks, and 20 replies, while the fourth reminder generated 243 opens and 7 replies. As a result, 21 calls were scheduled in the third week, marking a record outcome.

This case demonstrates that a systematic approach—thoughtful filtering of investors, domain warming, clear sequences comprising an initial email and three follow-ups with continuous metric analysis—enables stable increases in open rate, click rate, and conversion into meetings. The growth in scheduled calls from 14 to 21 over three weeks and the progressive improvement in engagement confirm the effectiveness of the described algorithm.

Conclusion

As a result of the conducted research, it has been demonstrated that an algorithmic approach to identifying relevant investors based on machine learning methods fundamentally alters the efficiency of the cold communication channel. In contrast to traditional manual research, which fails to achieve acceptable open and reply rates without warm introductions, the automated model sustains an open rate of 74–80% and achieves a reply rate of 10–17%, more than twice the average industry benchmarks. Meanwhile, the multistep follow-up strategy and rapid updating of investor data create a controlled feedback loop that enables iterative improvement of recommendation quality and conversion into calls: in the author's campaigns, 96 meetings were scheduled over one quarter without a single warm introduction.

The key success factor is the multilayer feature architecture combining structured deal registries, thematic embeddings, soft signals from public sources, and dynamic activity indicators. Such a combination enables the model to account for both historical connections within the investment graph and current shifts in fund focus, which is especially critical in a highly competitive environment for partner attention. Regular replenishment and updating of the training dataset, together with discipline in adhering to the cadence strategy, ensure algorithmic resilience to drift and preserve recommendation efficacy over the long term.

Practical implementation of the described algorithm encompasses not only the development of an investor ranking model but also the construction of a comprehensive infrastructure for large-scale mailings: from mail domain configuration and A/B testing of email templates to integration with meeting-scheduling tools and monitoring of key metrics. Taken together, this transforms fundraising from a lottery into a reproducible business process in which each stage, from initial contact selection to subsequent result analysis, is grounded in reliable data and continuous improvement.

The author's campaign results, clearly presented across three mailing waves, confirm that a systematic approach enables not only rapid growth in investor open and engagement rates but also steady increases in scheduled calls: from 14 to 21 over three weeks with fine-tuned scenarios and segmentation. This indicates that success in capital raising under modern conditions is determined not by the breadth of personal networks but by the quality of the algorithm and the rigor of operational discipline.

Thus, algorithmic investor identification via machine learning and disciplined multistep outreach transform the fundraising process into a controlled, repeatable procedure with clear metrics and opportunities for constant refinement. In an era of accessible data and cloud models, the primary determinant of outcome becomes not an elite network but a startup's ability to build and maintain a high-precision algorithm that places the most relevant investors at the top of the funnel.

References

1. Batrawy, B., & Cottle, C. (2024). *The Art of Cold Outreach*. Calendly. https://pages.calendly.com/rs/482-NMZ-854/images/The-Art-of-Cold-Outreach-EB.pdf

2. Campaign Monitor. (n.d.). *What are good open rates, CTRs, & CTORs for email campaigns?* Campaign Monitor. Retrieved May 15, 2025, from https://www.campaignmonitor.com/resources/knowledge-base/what-are-good-email-metrics/

3. Chatelaine, J. (2024, March 14). *The Complete 2024 Guide to Cold Email Metrics*. Quick Mail. https://quickmail.com/cold-email/metrics

4. Kaiser, M., & Kuckertz, A. (2024). Emotions and entrepreneurial finance: Analysis of venture capitalists' and business angels' digital footprints on Twitter. *International Entrepreneurship and Management Journal*, *21*(1). https://doi.org/10.1007/s11365-024-01018-9

5. McGee, C. (2024, August 29). *Average Cold Email Conversion Rate by Industry*. Focus Digital. https://focus-digital.co/cold-email-conversion-rate/

6. Piloterr. (n.d.). *Crunchbase Company Dataset*. Piloterr. Retrieved May 18, 2025, from https://www.piloterr.com/datasets/crunchbase-company

7. Send Grid. (n.d.). *2023 Email Marketing Benchmark Report*. Send Grid. Retrieved May 23, 2025, from https://sendgrid.com/en-us/resource/2023-Email-Marketing-Benchmark-Report?utm_source=chatgpt.com

8. Te, Y.-F., Wieland, M., Frey, M., Pyatigorskaya, A., Schiffer, P., & Grabner, H. (2023). Making it into a

successful series a funding: An analysis of Crunchbase and LinkedIn data. *The Journal of Finance and Data Science*, *9*, 100099. https://doi.org/10.1016/j.jfds.2023.100099

9. Thornhill, J. (2025, June 12). *Are machines smarter than venture capitalists?* Financial Times. https://www.ft.com/content/1cd7af37-2583-429e-b4d4-6d6a21108d75

10. Venture Capital Association. (2023). *Venture capital market study 2023*. Venture Capital Association. https://www.bvkap.de/files/content/Studien/Venture%20capital%20market%20study_23_Honold%20et.%20al.pdf

11. *What are B2B Cold Email Response Rates?* (2024, November 23). LinkedIn. https://www.linkedin.com/pulse/what-b2b-cold-email-response-rates-2024-study-pipeful-3tuke