Predicting Startup Funding using AI and the CFA

By

Yan Katcharovski, Andrew L. Maxwell

November 2024

1. ABSTRACT

Using 600 verbal startup pitches, we explore the application of Artificial Intelligence (AI) and the Critical Factor Assessment (CFA) to predict business angel funding outcomes. Contrary to prior research focusing on features derived from pitch metadata and voice recordings, we employ a Large Language Model (LLM) to synthesize key CFA factors, which assess startups across eight critical dimensions. These synthesized features are then used in machine learning models, such as Naive Bayes and Support Vector Machines, to predict investment decisions. Our findings reveal that specific factor combinations such as Features & Benefits, Readiness, and Financial Expectations consistently result in improved predictive capabilities, achieving accuracy of 79%, F1 score of 0.85 and average precision of 0.84. We demonstrate that models using synthesized CFA factors can effectively screen investment opportunities, offering a scalable and cost-effective method for early-stage funding evaluations.

2. INTRODUCTION

Entrepreneurial ventures are vital drivers of progress, creating jobs and boosting regional economic growth. High-growth ventures (often linked to technology innovation) are a subset of these ventures, often characterized by their focus on disruptive innovation, scalable markets and replicable business models to achieve growth and success objectives (Blank & Dorf, 2012). These high-growth ventures often raise external funding from investors to cover substantial costs associated with developing a product, establishing a new organization and developing a new market in advance of its ability to earn revenue from sales (Kerr et al., 2014). Investments in early-stage startups are motivated by potential financial returns and the opportunity for investors to support pioneering innovations and innovators (Wiltbank et al., 2009). The primary funding sources of these early-stage ventures, after personal and family/friends, are Business Angels: individuals who invest their own money in return for equity in the business (Kerr et al., 2014).

In addition to these Business Angels who have no direct personal relationship with the entrepreneur, funding can come from banks, grants, venture capitalists, and crowdfunding. The investment decision process of each of these sources of capital is quite different, with banks making secure loans and the government providing grants to support the development and growth of ventures aligned with specific program objectives. Venture Capitalists are fund managers who invest other people's money, usually in larger rounds and at a much later stage in the venture creation process (Wong et al., 2009). The relationship of these VCs with their investors and the funded entrepreneur creates several agency issues, resulting in a different investment decision process than for a Business Angel (Mason, 2008). These differences are often exacerbated by the fact that VCs often provide funding subsequent to Business Angels with a focus on scaling the business (Mason & Harrison, 2008).

Crowdfunding is an increasingly important source of non-traditional funding, often facilitated through a third-party platform, such as Kickstarter or GoFundMe. However, it is often linked to advance purchasing a new product or service rather than a direct investment. It is increasingly being seen as an alternative way of raising smaller amounts of capital from many individuals, without some of the regulatory issues faced by Business Angels. Perhaps the most important aspect of crowdfunding is that it provides startups with a level of market validation and early customer engagement that can be used as a signal to Business Angels and even VCs (Belleflamme et al., 2014). However, in most cases, people supporting crowdfunding campaigns do so because they see themselves (or people they know) as potential customers.

In contrast, Business Angels (BAs), investing their personal funds, often view their ability to offer mentorship, share valuable industry connections, and provide hands-on support as a more significant contribution. Business angel investors usually take on investment opportunities that are too risky or early for funding sources, such as banks and institutional venture capital firms, and often fill an extremely significant funding gap in the early days of a startup. In 2021, angel investors in the United States invested approximately $29.5 billion in over

69,060 ventures (Sohl, 2022), which accounts for 8-9% of total venture financing dollars. In comparison, VC firms invested approximately $315.6 billion across 13,050 deals (CB Insights, 2022). Although VCs account for a more significant percentage of total funds invested, BAs account for 84% (69,060 by angels vs. 17,054 by VC firms) of the total number of startup investment deals.

Given the importance of these early-stage, high-potential ventures in creating jobs, regional economic wealth creation and taking innovative products to market, the role and decision processes are of great interest to entrepreneurs and scholars alike, as well as those interested in seeing the impact of high growth ventures. These early-stage fund-seeking ventures face significant challenges in securing investment, as investors face the challenges of making a decision in the absence of a track record and market demand (Maxwell et al., 2011).

One of the biggest challenges facing early-stage startups (especially those run by inexperienced entrepreneurs) is that they do not understand the nature of the potential relationship with a Business Angel investor, how the Business Angel investor makes their decisions, and how they evaluate different forms of evidence. Indeed, there is strong evidence that Business Angels make poor decisions driven by inexperience, personal bias, and false preconceptions. Research by Daniel Kahneman, outlined in Thinking, Fast and Slow, demonstrates how human judgment is often affected by biases such as anchoring, availability heuristic and overconfidence, which make it even more challenging for the early-stage entrepreneur to attract investment. One approach to improve decision accuracy is to develop a structured decision approach (Kahneman, 2011). A better understanding of the actual investment decision process, and mainly the reasons specific opportunities are rejected, can help entrepreneurs learn and improve both their business and how they present evidence during an interaction with a Business Angel, improving the likelihood of securing investment (Cope, 2011).

In this research, we leveraged the research of Maxwell and Levesque to understand the Business Angels Investment Decision process better, using a range of AI tools that were not previously available. In their research, they used a widely deployed tool. The Critical Factor Assessment (CFA) (Maxwell & Lévesque, 2014) was used to predict early-stage rejection of an opportunity based on 8 Critical Factors (See 2.1). While the CFA tool has proven to be very accurate in predicting both the long-term success of a venture and which startups BAs will invest in, implementing the CFA suffers from both time (up to six weeks) and cost ($1395) constraints, requiring a team of trained experts to code and analyze evidence provided by the entrepreneur (sometimes directly and sometimes in response to BA questions). Recent rapid developments in Artificial Intelligence (AI) and Machine Learning (ML) offer a promising solution to overcome these limitations. They enhance objective assessment capabilities and reduce cognitive biases, such as overconfidence or anchoring, which often affect human evaluators and skew their judgment (Kahneman, 2011). Once an AI model is trained, it can provide cost-effective, consistent, and real-time processing capabilities for evaluations, enabling timely decision-making and feedback (Goodfellow et al., 2016).

While there have been some attempts to combine AI and investment decision-making frameworks, results are generally poor. In our paper, we build upon existing research in AI investment prediction by combining the modified CFA with the latest advancements in Large Language Models (LLMs) such as GPT. We investigated whether an LLM can accurately grade a startup pitch according to the 8 CFA factors, which were, in turn, used as synthesized features. Next, we trained custom AI models using the synthesized features on a dataset of startup pitches. The dataset used in the paper consisted of 600 Shark Tank pitches in the form of text transcripts, with their corresponding investment outcome (yes/no). The dataset used closely resembled Dr. Maxwell's research, which studied BA decision-making on the reality TV show Dragon's Den - a very similar format to the Shark Tank. We were further motivated to investigate BA investments as they account for the vast majority of investment deals. Furthermore, BAs play a critical role in the early-stage investment lifecycle of a startup, which is an extremely sensitive period crucial to their success.

This paper focuses on the following research questions:

1. *How accurately can an integrated CFA-AI model grade startup pitches following the 8 CFA factor format?*
2. *How accurately can an integrated CFA-AI model predict the investment decision making outcome of a Shark Tank pitch?*

3. LITERATURE REVIEW

2.1 Critical Factor Assessment (CFA)

The Critical Factor Assessment (CFA), developed in the 1980s, is an evaluation framework designed to analyze and score the commercial viability of a new venture across key factors critical to its success. It was developed by the Canadian Innovation Centre (CIC) in response to the need for a structured approach to assess innovative technologies and startup ventures during a period of rapid technological advancements (Canadian Innovation Centre, 1986).

At the time of the CFA's development, there was relatively little research on investor decision-making. Many aspects of the investment process were not well understood and were often viewed as "personal chemistry" or "gut feelings." Although venture capitalists often highlighted these elements as crucial, they remained largely underexplored by researchers (Hisrich & Jankowicz, 1990).

The original CFA model used by the CIC included 37 specific factors covering essential aspects such as market need, technical feasibility, financial viability, intellectual property protection, and management capability (Canadian Innovation Centre, 1986). While the 37 factors were generally effective in predicting commercial success, it was later found that they did not equally contribute to accurate predictions (Astebro, 2004) Astebro found that four categories of factors contributed the most while predicting the success of R&D projects: expected profitability, technological opportunity, development risk, and appropriability conditions (Astebro, 2004).

2.2 Dr. Andrew Maxwell's CFA

Joining the CIC in 2004, Dr. Andrew Maxwell began working with the CFA in the context of academic and industry application (A notable use case is the partnership with the Innovator's Assistance Program (IAP), mentioned in Daniel Kahneman's Thinking Fast and Slow). Leveraging data from the Canadian reality TV show "Dragons' Den", Maxwell sought to investigate whether the CFA would be effective at predicting the investment decision outcome of business angels on the show. Three trained CFA raters evaluated every business pitch across a grading rubric utilizing the critical factors. These factors included product adoption, product status, protectability, customer engagement, route to market, market potential, relevant experience, and financial model (Maxwell et al., 2011). Each of the factors were given a grade of A, B, or C, utilizing "+" or "-" to further differentiate the grades. Assigned grades were then converted to a numerical value from 0 to 10. A gap was introduced to the scoring when "jumping up" a grade letter, such as going from C+ to B-, to ensure a clear emphasis between investment readiness (Maxwell et al., 2011). Maxwell's findings confirmed that business angels use a heuristic called elimination-by-aspects (EBA) to filter out investment opportunities with fatal flaws. The study demonstrated that BAs rejected opportunities with fatal flaws (a grade of C) with 100 percent accuracy. The model achieved a general accuracy of 87.5 percent when predicting which ventures would advance beyond the initial selection stage (Maxwell et al., 2011). While the model demonstrated impressive predictive capabilities, the CIC struggled. The extensive data collection process required multiple trained evaluators working over multiple weeks. The time-consuming process of grading startups and the high costs associated with the evaluation make it challenging for organizations with limited staff and budget constraints to implement (Maxwell & Lévesque, 2014).

2.3 AI for Investment Decision Making

Artificial intelligence can significantly improve the efficiency and efficacy of evaluation frameworks and processes. Once properly trained and deployed, it can allow for more rapid and informed decision-making for startups (Goodfellow et al., 2016). Such models forego the cognitive bias of humans, as well as other factors such as fatigue and availability. AI models have the advantage of being able to continuously learn. As new data, discoveries, and advancements emerge, AI models can be updated and trained to enhance predictive accuracy and adapt to evolving market conditions (LeCun, Bengio, & Hinton, 2015). This adaptability is particularly beneficial for startup evaluation, where the volume and complexity of data can overwhelm human evaluators.

In recent years, research applying AI for investment decision making using the Shark Tank television show as a dataset for predicting startup investment outcomes has been on the rise. While there are many drawbacks to using a data source such as Shark Tank (discussed later in the paper), the accessibility to public data, data privacy, and standardization across all episodes make it an appealing source for researchers.

"Shark Tank Deal Prediction Using Machine Learning Techniques" by Deodhar, Bhatkar, and Dixit (2024) utilized machine learning techniques to forecast the outcomes of deals made on Shark Tank India. The dataset included 117 instances collected from the show's first season, featuring variables such as investment amounts, startup valuations, entrepreneur characteristics, team composition, equity offering, product category, and the presence of specific sharks on each episode (as guests often rotate and introduce new decision-making variables). The researchers utilized multiple machine learning models, such as Artificial Neural Networks (ANN), Random Forests, Decision Trees, and Logistic Regression, to predict whether a deal would occur. The study revealed that ANN and Random Forest performed the best, achieving an accuracy of 67% in predicting successful deals. Techniques such as Random Forest and ensemble learning encapsulated variability in the features pertaining to the pitches, while ANN was deployed to capture nonlinear relationships among the variables. While the study showed a general viability of utilizing machine learning models on the Shark Tank dataset, it displayed a few limitations. The small dataset of 117 pitches has a significant impact on AI model performance, especially for models like Logistic Regression and Decision Trees that rely on larger datasets. These dataset sizes pose the risk of training an overfitted model.

Furthermore, the study used general metadata describing the characteristics of the pitch (mentioned above), rather than information from the pitch itself. This limited the model's ability to provide comprehensive analysis, reducing the depth of insights that could be derived. To address these challenges, richer datasets incorporating more detailed qualitative and behavioral factors are needed.

A similar study, "Shark Tank Deal Prediction: Dataset and Computational Model" by Sherk, Tran, and Nguyen (2019), used a Shark Tank dataset with features such as product category, team demographics, valuation, equity offered, and the presence of specific sharks to predict investment outcomes. The researchers achieved a deal/no deal predictive accuracy of 62.5% using logistic regression and neural networks. The study highlighted the importance of both financial and emotional appeal in the verbally spoken pitch in securing investments.

A review of these and similar papers concluded in consistent results in terms of accuracy, prompting the need to investigate the possibility of acquiring and analyzing more complex data sources (compared pitch metadata, i.e., ask amount, team composition, startup industry, etc.) that would yield richer input features.

Another relevant paper, "Blood in the Water: An Abductive Approach to Startup Valuation on ABC's Shark Tank" by Lavanchy, Reichert, and Joshi (2022), investigated startup valuations and investor decision-making on the show. Decision tree models were utilized to analyze how initial equity offers and requested valuations influenced the likelihood of securing investment. The researchers concluded that startups offering more than 15% equity were significantly more likely to secure deals, with the decision tree model predicting deal success with an accuracy of 70%.

When it comes to using the actual verbal pitch content, similarly to Dr. Maxwell's research, the study "Pitch Perfect: Predicting Startup Funding Success Based on Shark Tank Audio" by Raghvendra, Wood, and Xiao (2017) stood out. The researchers focused on analyzing audio features from Shark Tank pitches to predict whether a startup would secure funding. They utilized a hybrid CNN-LSTM model to capture both pitch delivery's sequential and contextual aspects, extracting features such as Mel-Frequency Cepstral Coefficients (MFCC), prosodic elements (e.g., pitch and intensity), and other emergent speech characteristics. The hybrid model achieved a validation accuracy of 68% and test accuracy of 66%, demonstrating a moderate feasibility of using speech data to gauge persuasiveness in startup pitches. The researchers encountered several challenges, including noisy audio data due to the dramatic nature of Shark Tank and limitations in dataset size, which consisted of 509 audio segments. Despite the implementation complexity and conservative predictive accuracy, the study showcased how advanced neural network architectures could model complex persuasive cues in spoken pitches.

4. METHODOLOGY

"Shark Tank" is a US-based reality TV show in which entrepreneurs pitch their startups to a panel of Business Angel investors, known as the "Sharks". By pitching their startup, entrepreneurs seek to raise funding for in exchange for company equity. The Sharks listen to the pitches, ask questions, and negotiate deals if they see potential in the business. The format simulates real-world investor-entrepreneur interactions and involves real Business Angels investing their own money. Conversations between the entrepreneurs and Sharks provide insight into how investment decisions are made based on factors such as product potential, business model, and financials.

The primary data source was the reality TV show "Shark Tank," which followed a similar methodology used by Maxwell with the television show, "Dragons' Den." In Maxwell’s 2011 paper, “Business Angel Early Stage Decision Making,” he addressed the validity of TV show data as a representation of real-life investor decision-making, noting that the interactions on Dragons’ Den resembled real-life circumstances. This provided a foundation for using similar television show data, such as Shark Tank, as a credible source for studying investor behaviors and decision processes.

In total, 1153 records of pitch data were collected from a publicly available CSV project called “Shark Tank US dataset” on the AI web-community Kaggle, that represented a diverse array of industries, pitch styles, and investment outcomes. While this dataset contained 53 fields of data, including startup industry, investment by each Shark, season number and date, etc., we only utilized the fields containing Season, Company Name, Deal (containing 0 or 1 for no deal/deal), Ask Amount and Ask Equity.

The next phase entailed the collection of subtitles for each pitch. We were able to acquire 600 subtitle transcripts from publicly available repositories such as Subdl and OpenSubtitles.org. The dataset was then cleaned to address missing values and standardize formats ensuring suitability for AI model training.

4.1.SYNTHESIZING CFA FEATURES

We built a back-end web-application capable of communicating with OpenAI's GPT API, leveraging Function Calling to accept and return structured text and number data, as well as Fine-Tuning to train and customize GPT's response performance for our analysis of Shark Tank pitches. The web application, called Expitch.com, served 600 requests to the API which included the full transcript of every Shark Tank pitch, as well as eight prompt-engineered CFA rubrics. The CFA rubrics were built using the Canadian Innovation Center's (CIC) CFA snapshot tool (Table 1) and grade conversion table (Table 2). The numerical jump between C+ to B-, and B+ to A- reflects the significant difference between each grade category (Maxwell, 2011), as discussed earlier in the paper.

Factor	**Criteria**	**Evaluation For Grade**
1. Features & Benefits	Performance Advantages	Does the proposed product offer advantages over currently deployed solutions?
	Benefits & Costs	Are there substantial advantages, benefits, or competitive performance at a competitive cost?
	Customer Demands	Can the solution meet customer demands and expectations?
2. Readiness	Product Delivery	How close is the startup to delivering completed products/services to first customers?
	Milestones Achieved	Have product development milestones been completed, covering technology, manufacturing, and supply chain?
	Validation Tests	Are there sales or beta tests validating readiness?
3. Barrier To Entry	Uniqueness	What unique/patentable features represent a barrier to entry for competitors?
	Proprietary Technology	Are there patents or proprietary tech not easily replicated?
	Brand & Features	Are there unique features or brand elements providing significant competitive barriers?
4. Adoption	Market Purchase	Can the startup demonstrate customer intent to buy the product/service?
	Customer Commitment	Have customers been involved in development and committed to purchasing?
	Market Validation	Is there evidence or independent market validation indicating customer interest?
5. Supply Chain	Barrier Confirmation	Are there no barriers concerning the supply chain/distribution channel?

	Supplier Engagement	Are suppliers and channel partners engaged, with commitments upon readiness?
	Partner Identification	Have potential supply chain partners been identified or approached?
6. Market Size	Market Potential	Is there evidence supporting large market potential and achievable market share?
	Market Validation	Is there some evidence of a large potential market and sales targets?
	Revenue Forecasting	Is there evidence making revenue forecasting possible or is it challenging?
7.Entrepreneur Experience	Relevant Experience	Does the team possess significant relevant experience?
	Business Experience	Does the team have significant but non-direct business experience?
	Technical Limitation	Does the team primarily have technical experience without evidence of management capability?
8.Financial Expectations	Cash-Flow Confidence	Does the financial projection indicate a high degree of confidence in achieving cash-flow neutrality?
	Projection Reliability	Is there a reasonable confidence level in achieving cash-flow neutrality, either directly or through fundraising?
	Negative Position	Does the projection lack detail or show negative cash flow with limited confidence in fundraising?

Table 1 – CIC Snapshot Critical Factor Assessment

GRADE	SCORE	ADJUSTED SCORE
A+	10	80
A	9	72
A-	8	64
B+	6	48
B	5	40
B-	4	32
C+	2	16
C	1	8
C-	0	0
N/A	0	0

Table 2 – CIC Grade Conversion Table

The returned CFA grades were organized on the 600-record CSV to include the actual investment outcome. All data was codified and converted to numerical format, ready for AI model training.

4.2.AI MODEL ASSESSMENT OF FEATURES

A multi-model machine learning evaluation pipeline was developed to predict the target variable "Deal" using a structured methodology. Data containing the newly synthesized CFA features was imported from the main CSV file. Standard data cleaning practices were implemented such as dropping rows with empty columns and encoding categorical targets.

Feature engineering involves generating all possible combinations of specified feature subsets to explore their predictive efficacy. The dataset was then partitioned into training and testing sets using stratified sampling to maintain class distribution. A diverse array of classification algorithms, including Logistic Regression, Neural Networks, Decision Trees, Ensemble Methods (e.g., Random Forest, Gradient Boosting, AdaBoost), Support Vector Machines, Naive Bayes variants, XGBoost, and Linear Discriminant Analysis, were systematically evaluated. Each model underwent hyperparameter tuning through RandomizedSearchCV within a cross-validation framework to optimize performance metrics such as accuracy, F1 score, specificity, and a custom-balanced score. Random Undersampling was integrated into the preprocessing pipeline to address the class imbalance. The evaluation was parallelized using joblib to enhance computational efficiency. At the end, the top-performing models based on the balanced score were retrained on the entire training set and assessed on the test set to validate their generalizability. This strategy was implemented to conserve data efficiency during the validation and testing phase, and ensure all data is being used for training. Final results, including detailed performance metrics and visualizations, were saved and analyzed to identify the most effective models and feature combinations.

5. RESULTS

While non-AI implementation of the Dr. Maxwell's CFA underscores the importance of evaluating all 8 factors, our results concluded that specific feature combinations outperformed the rest. Specifically, features such as Features & Benefits, Readiness, Barrier to Entry, Supply Chain, Entrepreneur Experience, and Financial Expectations consistently appeared in the top-performing models, underscoring their significant contribution to predicting the target variable "Deal." In contrast, the Adoption factor did not appear in any models, and Market Size only appeared twice, indicating that their predictive impact is none or low (within the current modelling framework). These patterns are to be addressed in the discussion section. A custom Balanced Score was introduced to assess each model's performance, calculated by taking the average of the F1 Score, Accuracy, and Specificity. This score was used during the cross-validated stage to compare each model's ability to correctly identify positive (deal) and negative (no deal) classes, thereby addressing class imbalance and ensuring a comprehensive performance assessment.

Among the models evaluated (Table 3) Naive Bayes emerged as the top model, achieving the highest test accuracy of 79% with the feature combination (Feature & Benefits, Barrier To Entry, Entrepreneur Experience, Financial Expectations, and the Total sum of all factors). This model demonstrated a balanced performance (when it comes to performance for both 0 and 1 classes) with a test F1 Score of 0.85, Specificity of 0.63, ROC AUC of 0.74, Average Precision of 0.84, and Recall of 0.87, highlighting its effectiveness in accurately predicting deals while maintaining adequate specificity. While most of the reviewed papers did not focus on Specificity, we believed it was an important metric to analyze, as predicting true negatives (which startups will NOT receive investment), is a major contributor to the EBA hypothesis presented by Dr. Maxwell's research. Support Vector Machines (SVM) also stood out, attaining a test accuracy of 74.44% and a superior test F1 Score of 0.81 using the feature combination (Feature & Benefits, Readiness, Barrier to Entry, Entrepreneur Experience). SVM also achieved a Specificity of 0.63, ROC AUC of 0.70, Average Precision of 0.79, and Recall of 0.80, demonstrating its ability to balance precision and recall effectively. Logistic Regression matched SVM's test accuracy of 74.44%, with a test F1 Score of 0.75, Specificity of 0.69, ROC AUC of 0.72, Average Precision of 0.81, and Recall of 0.74 using the feature combination (Feature & Benefits, Barrier To Entry, Financial Expectations, Entrepreneur Experience). This model showcased improved specificity compared to SVM, indicating a better ability to identify negative classes correctly. These top three models benefited from iterative feature selection and preprocessing techniques, including Random Undersampling and the adoption of binary validation metrics, which enhanced their specificity and overall performance. It is worth discussing how to correctly determine which model is considered "best", depending on the circumstance and desired sensitivity to the Positive vs. Negative class.

Model	Factor Combination	Test Accuracy	Test F1 Score	Test Specificity	Test AUC	Test Average Precision	Test Recall
Naive Bayes	(1, 3, 7, 8, Total)	79%	0.85	0.63	0.74	0.84	0.87
Naive Bayes	(1, 2, 5, 6, 7)	77.8%	0.77	0.6	0.74	0.84	0.78
SVM	(1, 2, 3, 7)	74.4%	0.81	0.63	0.7	0.79	0.8
Logistic Regression	(1, 3, 5, 7)	74.4%	0.75	0.69	0.72	0.81	0.74
Neural Network	(1, 5, 8)	74.4%	0.75		0.74	0.82	
Voting Ensemble	(1, 2, 5, 7, 8)	74.4%	0.74	0.69	0.68	0.78	
Naive Bayes	(3, 5, 7, 8)	74.0%	0.81	0.63	0.8	0.87	0.8
Naive Bayes	(1, 6, 7, 8)	73.0%	0.79	0.7	0.72	0.82	0.75
Naive Bayes	(1, 3, 5, 8)	71.0%	0.76	0.73	0.75	0.84	0.7

Table 3 – Model Results, for factor enumeration refer to Table 1

6. DISCUSSION

Our study aimed to evaluate the effectiveness of an integrated Critical Factor Assessment (CFA)-AI model in grading startup pitches and predicting investment decisions, using "Shark

Tank" pitches as a dataset. The results demonstrate that our AI model, which leverages a Large Language Model (LLM) to extract CFA features and combines them with custom machine learning algorithms, achieves a comparable predictive accuracy and, in some respects, is superior to traditional methods like Andrew Maxwell's CFA.

6.1.COMPARISON TO TRADITIONAL CFA

Andrew Maxwell's CFA method has been shown to predict early-stage investment decisions with an accuracy of 87.5% (Maxwell et al., 2011). Our integrated CFA-AI model achieved a test accuracy of up to 79% using Naive Bayes classifiers, with other models, such as Support Vector Machines and Logistic Regression, achieving accuracies of 74.44%. While our CFA-AI model performs slightly worse than Maxwell's human-graded method, it is important to note the differences in approach and data availability. In this study's dataset, verbal pitches spanned 3-5 minutes on average, whereas Maxwell had access to richer entrepreneur-investor conversations spanning 30-60 minutes.

In contrast to Maxwell's model's proficiency in predicting negative (no deal) classes using fatal flaw factors, our CFA-AI model performed better when predicting the positive (deal) class. Whereas Maxwell's model performed best when using all eight CFA factors, our strongest models utilized three to six. Additionally, while Maxwell's model aggregated the scoring of 3 individual CFA raters, our model considered only a single output from the LLM, which could introduce variance to the results.

Interestingly, the Adoption factor did not appear as a significant predictor in our models. This contrasts with Maxwell's findings, where the Adoption factor was critical in early-stage investment decisions. A possible explanation is that pitches featured on "Shark Tank" typically have already achieved some level of market presence or customer traction, which is a prerequisite for being selected for the show. Consequently, there may be less variance in the Adoption factor among these pitches, reducing its predictive power in our dataset. Further research is needed to investigate this phenomenon

6.2.COMPARISON TO OTHER SHARK TANK STUDIES

Similar studies using Shark Tank data discussed in this paper have reported predictive accuracies ranging from 62.5-70%. While some of the techniques used in this paper resemble previous research, our unique CFA-AI model offered notable advantages.

For instance, Deodhar et al. (2024) utilized machine learning models like Artificial Neural Networks (ANNs) and Random Forests on a dataset of 117 pitches from Shark Tank India, achieving an accuracy of 67%. Sherk et al. (2019) applied logistic regression and neural networks to a Shark Tank dataset, achieving a predictive accuracy of 62.5%. Similarly, Lavanchy et al. (2022) employed decision tree models to analyze startup valuations and investor decision-making, attaining an accuracy of 70%. Raghvendra et al.'s (2017) paper stood out due to their focus on analyzing audio features from "Shark Tank" pitches using a hybrid CNN-LSTM model, achieving a validation accuracy of 68% and a test accuracy of 66%.

Our models outperformed these studies, achieving predictive accuracy of up to 79%. Some of the aforementioned studies relied on limited input features extracted from show metadata when training their models, such as financial metrics, team demographics, and categorical variables, which lack informational depth and content. In contrast, our model leverages the full textual content of pitches, analyzed through the lens of the eight CFA factors.

6.3.VALIDITY OF SYNTHESIZING CFA FEATURES USING LLMS

One of the key contributions of our research is demonstrating the validity of using an LLM to extract CFA features from startup pitches. While LLMs such as GPT are trained on a vast dataset and excel at conversational text generation, fine-tuning using CFA rubrics proved to be successful (by proxy of overall model success, specific validation to be considered in future research).

6.4.RESEARCH LIMITATIONS

Our study has several limitations that should be acknowledged. First, the use of "Shark Tank" data, while convenient and rich, may not fully represent real-world investment scenarios. The show is edited for entertainment purposes, and pitches may not reflect the depth and breadth of information typically provided in investor presentations. While this was addressed in Maxwell's 2011 research, he was able to access the full conversation records, compared to our shortened TV-version of the pitch (Maxwell et al, 2011). Additionally, our dataset, although larger than those used in some previous studies, is still relatively small for training robust AI models. Another limitation is the class imbalance in our dataset, with approximately 66% of pitches receiving investments and 34% not. Although we applied balancing techniques such as Random Undersampling, this imbalance may still have influenced our models' performance, particularly in predicting negative investment outcomes.

We also experimented with various machine learning models and approaches, including vectorizing the entire transcript text as features. However, models using vectorized text alone yielded poor results. This suggests that while textual data is rich, extracting meaningful predictive features requires careful feature engineering, such as our approach of using LLM-extracted CFA factors.

6.5.RESEARCH QUESTIONS ADDRESSED

How accurately can an integrated CFA-AI model grade startup pitches following the 8 CFA factor format?

Our integrated CFA-AI model, utilizing an LLM to extract CFA factors, was able to successfully grade startup pitches. The LLM effectively interprets pitch transcripts and assigns grades consistent with the CFA criteria. However, we were not able to validate whether the LLM's synthesis of the grades was accurate relative to expert human grading. This will be addressed in our future research. In this paper, we are able to make claims about the LLM's success in synthesizing features only by proxy of the overall LLM-Naive Bayes success in predicting Shark Tank investment outcomes.

How accurately can an integrated CFA-AI model predict the investment decision-making outcome of a "Shark Tank" pitch?

The integrated CFA-AI model achieved a predictive accuracy of up to 79%, outperforming previous studies using Shark Tank data. Metrics such as F1 score, ROC AUC, and recall, were used, as well as a Balanced Score, taking into account Accuracy, F1 score and Specificity while validating and testing the models, indicating its effectiveness in predicting investment outcomes based on extracted CFA features. We will continue to further improve our model with future LLM improvements and access to new data sources.

6.6.PRACTICAL APPLICATIONS

The findings of our research have significant practical implications:

- **For Entrepreneurs and Educators:** A deployed and publicly accessible model based on our CFA-AI research can provide real time feedback on pitches across eight critical business areas (note that our model did not make use of Adoption in the top performing models). Such a tool (see expitch.com) can guide them in refining their business plans and presentations to increase the likelihood of securing investment.
- **For Investors:** The AI model offers a rapid and consistent method for preliminary screening of investment opportunities. While sophisticated and experienced investors would likely prefer to rely on their own judgement, our model can be used by angel groups, individuals looking to get into angel investment, or associates and analysts of investment firms.

7. FUTURE RESEARCH

While our study demonstrates the potential of integrating Large Language Models (LLMs) with the Critical Factor Assessment (CFA) framework to predict investment decisions, there are several areas of improvement and exploration which could enhance our research further.

7.1.Generating Useful Insights for Entrepreneurs

While not discussed in this paper, our CFA-AI tool (Expitch), is able to generate written analysis based on the CFA grading rubric. Output results identify strengths and weaknesses in the pitch while also providing a justification for the given grade. Further to the written evaluation, the tool also generates a recommendation, which follows the CFA rubric and provides insights for reaching a grade of A+ in the given factor.

Future research could involve longitudinal experiments where entrepreneurs utilize and implement AI-generated feedback, followed by assessments of changes in investment readiness and success rates. Qualitative surveys and interviews with entrepreneurs could extract perceived usefulness and actionable nature of the model's insights.

7.2.Assisting Business Angels in Informed Decision-Making

While BAs were consulted during the development of our model, an opportunity exists to research whether a commercialized implementation of our tool would tangibly help BAs. While making accurate and informed decisions as to which startup is likely to succeed or fail is at the top of every investor's agenda, there are various use cases that span beyond.

Since human judgment is often subject to cognitive biases such as overconfidence and anchoring (Kahneman, 2011), incorporating AI assessments could mitigate these effects. To validate this hypothesis, the integration of our tool as an assistant during the BA investment decision making process and measuring the impact on outcomes could yield promising results.

7.3.Validating LLM Accuracy through Expert Comparison

An immediate next step proposed beyond this paper includes a comprehensive cross-validation of the LLM's synthesized CFA features with expert CFA-graded reports. This would require the acquisition of a dataset of startup evaluations by a panel of raters with known outcomes. Data used in Astebro's research (2004) described a dataset from the IAP which also included the commercialization success of the R&D projects a few years past CFA evaluation. Such a dataset could be extremely valuable in analyzing synthesized CFA grades as well as the predictive capabilities of a startup's future success or failure.

7.4.Refinement of Feature Extraction Techniques

While our use of an LLM to synthesize CFA features has shown promise, the implementation of advanced techniques such as Natural Language Processing (NLP) could further enhance input features. Techniques such as sentiment analysis, topic modeling, and semantic role labeling could be implemented, allowing the model to capture emotional and contextual cues that influence investment decisions. While this research utilized OpenAI's tools, future research would experiment with additional LLMs such as Google's Gemini and Anthropic's Claude.

7.5. Investigation of Underrepresented Factors

While Dr. Maxwell's model relied on the use of all eight critical factors, our AI model failed to use Adoption in any top performing results. While there could be a variety of reasons for this, such as poor feature extraction by the LLM, further research is required. One plausible explanation is that most startups appearing on Shark Tank have already reached a certain level of adoption (as a prerequisite to be selected for the show).

Analyzing datasets where startups are at varying stages of customer adoption could provide insights into the factor's predictive power.

REFERENCES

Astebro, T. (2004). Key success factors for technological entrepreneurs' R&D projects. IEEE Transactions on Engineering Management, 51(3), 314-321.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. Journal of Business Venturing, 29(5), 585–609.

Blank, S., & Dorf, B. (2012). The Startup Owner's Manual: The Step-by-Step Guide for Building a Great Company. Pescadero, CA: K & S Ranch Publishers.

CB Insights. (2022). State of Venture 2022 Report. Retrieved from CB Insights website.

Canadian Innovation Centre. (1986). Critical Factors Assessment. Retrieved October 2nd, 2024, from https://innovationcentre.ca/services/services-for-entrepreneurs/critical-factors-assessment/

Deodhar, O., Bhatkar, S., & Dixit, P. (2024). Shark Tank Deal Prediction Using Machine Learning Techniques. Journal of Business Research.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Kahneman, D. (2011). Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. Journal of Economic Perspectives, 28(3), 25–48. doi:10.1257/jep.28.3.25

Lavanchy, M., Reichert, P., & Joshi, A. (2022). Blood in the Water: An Abductive Approach to Startup Valuation on ABC's Shark Tank. Journal of Business Venturing.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Mason, C. M., & Harrison, R. T. (2008). Measuring business angel investment activity in the United Kingdom: A review of potential data sources. Venture Capital, 10(4), 309–330.

Maxwell, A. L., Jeffrey, S. A., & Lévesque, M. (2011). Business angel early stage decision making. Journal of Business Venturing, 26(2), 212–225.

Maxwell, A. L., & Lévesque, M. (2014). Technology venture assessment in practice: critical factor assessment and venture grading. R&D Management, 44(2), 176-188.

Raghvendra, K., Wood, M., & Xiao, L. (2017). Pitch Perfect: Predicting Startup Funding Success Based on Shark Tank Audio. Journal of Business Research.

Sherk, T., Tran, M. T., & Nguyen, T. V. (2019). Shark Tank Deal Prediction: Dataset and Computational Model. Journal of Business Research.

Sohl, J. E. (2022). The Angel Market in 2021: Angels Remain Resilient in the Face of Pandemic Challenges. Center for Venture Research, University of New Hampshire.

Wiltbank, R., Read, S., Dew, N., & Sarasvathy, S. D. (2009). Prediction and control under uncertainty: Outcomes in angel investing. Journal of Business Venturing, 24(2), 116–133. doi:10.1016/j.jbusvent.2007.11.004.

Wong, A., Bhatia, M., & Freeman, Z. (2009). Angel finance: The other venture capital. Strategic Change, 18(7-8), 221–230.

Cope, J. (2011). Entrepreneurial learning from failure: An interpretative phenomenological analysis. *Journal of Business Venturing*, 26(6), 604-623.

Hisrich, R. D., & Jankowicz, A. D. (1990). Intuition in venture capital decisions: An exploratory study using a new technique. *Journal of Business Venturing*, 5(1), 49-62.

Mason, C. M. (2008). *The real venture capitalists: A review of research on business angels*. Hunter Center for Entrepreneurship, University of Strathclyde.