

AI Startup Idea Validator

Krishna Rao P.¹, Aparna B.², Sravani K.³, Yasaswini S.⁴

¹Assistant professor, ^{2,3,4}Student, Department of Computer Science and Engineering,
Geethanjali College of Engineering and Technology (Autonomous), Hyderabad, Telangana,
India.

E-mail: ²22r11a05k3@gcet.edu.in

Abstract

Start-up validation at the initial stage demands a combination of business feasibility assessment, market environment, competitive positioning, technical feasibility, and financial sustainability, which is hard to accomplish manually. The proposed paper discusses a startup evaluation framework that combines Large Language Model (LLM) reasoning with business and financial analytics. The startup inputs such as problem description, target market, solution, value proposition, and business model are checked using the Flask backend and converted into structured prompts for LLM. The evaluation process will cover demand in the market, competition, uniqueness, scalability, feasibility, risk, and yield criterion-level results, viability ranking, and advice on how to improve. The finance module will handle income, profit margins, cash flow, burn rate, financing, leverage, cost per acquired customer, and life-time value to determine the financial health and viability of the start-up business. Persistence using MongoDB will allow for comparison with past evaluations.

Keywords: Large Language Model Reasoning, Startup Viability Assessment, Business-Contextual Analysis, Financial Sustainability Modeling, Automated Decision Support, Prompt-Structured Evaluation, Persistent Evaluation Analytics.

1. Introduction

The AI Startup Idea Validator based on Large Language Models (LLM) is a decision-support system meant to help entrepreneurs assess their startups in the early stage before investing a considerable amount of time and money. Artificial Intelligence (AI) systems have been used more often to perform business analysis, evaluate the market, and make decisions, while the large language models allow for processing complex business descriptions

in context. Nevertheless, the problem of assessing startups remains uncertain because of the need to analyze the market demand, competitive situation, scalability, feasibility, appropriateness, and risks.

Traditional methods for validation often rely on manual research, expert advice, and subjective opinions, making the initial validation cumbersome and non-standardized. Analysis of startups using an LLM is one way to make the initial validation automated through interpreting text and creating business insights from it. Nevertheless, a successful validation system cannot only create unlimited amounts of texts; the evaluation parameters should be standardized.

This system will address the above-stated criteria by collecting critical start-up information using a web interface and processing it using a structured LLM evaluation pipeline. The system will evaluate different viability parameters, calculate the overall viability score, consider any existing financial indicators, identify significant risks and produce suggestions for how to improve the idea. Collected information and produced reports will be saved in the MongoDB database and can be later accessed via the evaluation history interface. In addition, the system will consider financial indicators like revenue, cash flow, burn rate, runway, customer acquisition costs, and customer lifetime value.

To support objective evaluation, the system could be benchmarked with real-life Kickstarter campaign data that includes textual information about the campaigns as well as related parameters [11]. This would help in evaluating the consistency and effectiveness of the LLM-based startup evaluations with the actual results from the campaigns. In this way, the proposed system integrates contextual LLM evaluation, multi-parameter scoring, financial evaluation, and risk-based decision-making into one startup evaluation system.

2. Related Works

Recent research on startup has been observed that recent advances in the field of startup evaluation have largely moved away from traditional judgmental assessment of startups towards computational-based assessments using data. In machine learning, it has been seen that the features of successful ventures can be predicted using the techniques of machine learning from structured data about startups [2]. Market-based predictions have used internal venture features along with market-related conditions to predict startup performance and indicate that startup performance is affected by multiple factors [4]. Various studies have used other aspects of

startups such as the technology involved and venture capitalist related features for predicting startup success, and various dimensionality reduction techniques have been explored to represent startup features [6], [8].

Natural language information serves as yet another channel to analyze the early-stage businesses. It has been shown that through content analysis one could use natural language information of startups, along with cohort attributes, to generate meaningful cues about success of the ventures [5]. Later studies further developed this area using the idea of generative AI. The capacity of language models to analyze startups and produce business-level judgments has been analyzed in context of early-stage investments and survival prediction [1], [7]. Reasoning-based methods were applied to memory enhancement and multi-step reasoning to solve the problem of startup evaluation, underscoring the significance of systematic reasoning in this area [3].

Nevertheless, the use of generative AI poses difficulties related to hallucination, reasonability of the results produced, and interpretation of the produced entrepreneurial recommendations [9]. As a result, the approaches to AI decision-making have turned their attention to integration in order to provide startup-success prediction and risks evaluation by means of considering the significance of using several data sources and analysis methods [10]. In any case, it is possible to claim that all current approaches are oriented towards either predictive modeling of startup success or using LLMs for entrepreneurial analysis, which leaves room for an integrated approach to be developed.

3. Proposed Methodology

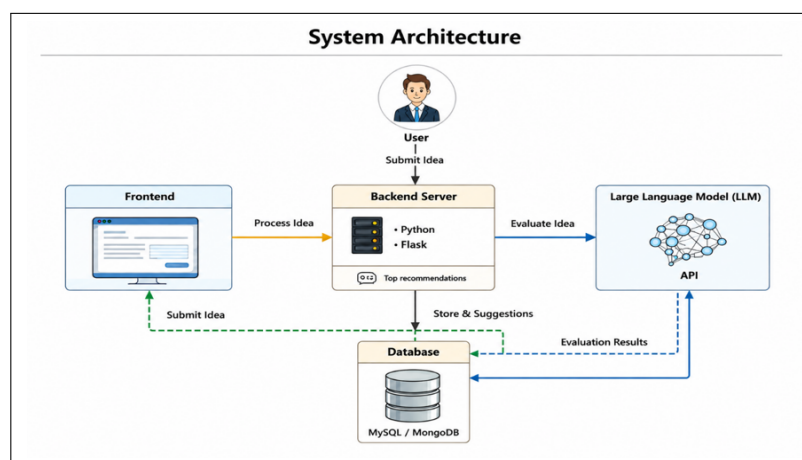


Figure 1. LLM-based startup validation architecture

The proposed AI Startup Idea Validator operates based on a pipeline processing methodology which consists in the translation of the user-supplied data related to the startups into a multidimensional viability assessment. As shown as Figure 1, the proposed approach encompasses input pre-processing, dataset-driven benchmarking, structured prompts design, LLM evaluation, viability rating, financial evaluation, risk assessment, recommendations provision, and persistent report storage.

3.1 Startup Input and Preprocessing

The system first takes in four fundamental startup attributes via the web portal: the problem, target audience, solution to the problem, and the business model. The above information is what is needed for the evaluation process. The Flask back-end ensures that the information provided is validated and then structured before sending it to the LLM as a request. This process standardizes the inputs from different startups so that the evaluation process becomes less ambiguous. The validated information is then sent to the prompt generation stage, as shown in Figure 2.

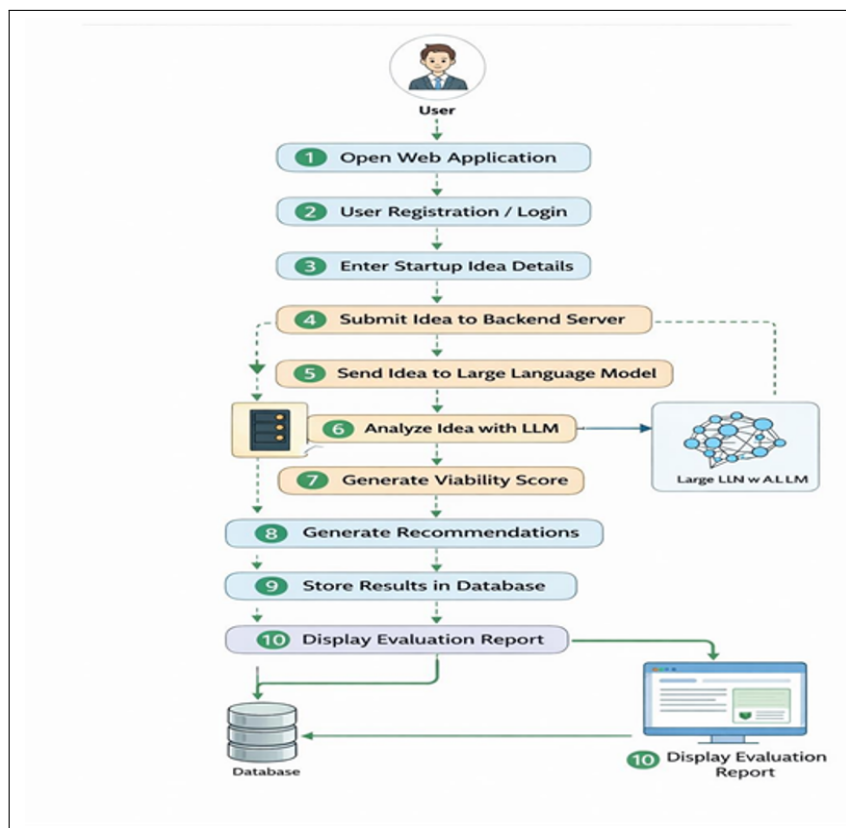


Figure 2. Operational workflow of the LLM-based startup idea validation system

3.2 Dataset Description

Kickstarter 2018 (Pre-processed Dataset) [11] serves as a benchmark dataset to evaluate the proposed approach of using an LLM to assess startups. This dataset comprises 191,724 campaigns with both text and structured features such as campaign name, blurb or description, funding data, campaign duration, and outcome of success or failure. The campaign description serves as the input in natural language form while the campaign outcome is kept as the evaluation reference.

To perform the classification evaluation using the quantitative method, the outcomes of the campaigns are categorized into three reference classes based on the decision classes generated by the model: Recommended, Moderate, and Not Recommended. Successful campaigns are considered to belong to the Recommended category. Unsuccessful campaigns are further divided depending on the achievement of the predetermined target funding level: those that reach more than 50% of the target funding are labeled as Moderate, while the rest are classified as Not Recommended. Data containing missing information about the campaign outcome is disregarded from the classification evaluation process.

The original outcome of the campaign is translated to the decision label space needed for the proposed assessment model. The successful campaign is mapped to Recommended while the unsuccessful campaign is mapped to Not Recommended. However, since the original outcome has a binary nature, the Moderate label cannot be taken from the original outcome and hence does not have an independently created ground-truth label without the definition of an intermediate-outcome criterion.

Before partitioning, any data that has missing or unusable required variables is excluded. Subsequently, the remaining data is split into two groups: 80 % for the development/reference purpose and 20 % for the held-out test dataset, maintaining the proportionate class of the target outcome variable. The LLM does not undergo training and finetuning on the dataset. The assessment generated by processing the description from the held-out dataset is then compared to the actual result obtained.

Table 1. Dataset specifications

Parameter	Specification
Dataset	Kickstarter 2018 (Pre-Processed Data)
Source	Zenodo
Campaign Records	191,724
Primary Textual Field	Campaign blurb/description
Supporting Attributes	Campaign name, funding date, funding period, funding amount, success status
Evaluation Target	Campaign success status
Dataset Role	Benchmark evaluation
LLM Training/Fine-Tuning	Not performed
Development/Reference Subset	80%
Held-Out Test Subset	20%
Split Strategy	Stratified random partition
Text Utilization	Natural-language startup/idea input
Preprocessing	Removal of incomplete/unusable records

3.3 Structured Prompt Construction and LLM Evaluation

The validated representation of the startup is transformed into an organized prompt that has the problem statement, target market, solution offered, and business model. The prompt dictates the evaluation criteria to be used as well as the response structure to be provided by the LLM such that the LLM does not produce unrestricted narratives but rather focuses on startup assessment fields. The prompt created is sent through the Flask backend to the set up LLM API.

The response received from the LLM is parsed in order to retrieve the respective assessment scores as well as comments. These scores are then normalized against a common scale prior to being scored and financially processed. Therefore, the LLM acts as the contextual assessment layer, whereas scoring and financial processing interpret its output.

3.4 Multi-Factor Viability Assessment

Normalization of the extracted scores for the LLM assessment is done, based on a common scale, while the scores are aggregated based on the predefined criteria. Calculation of the total viability score is based on the weighted aggregation of the criteria, as expressed in the formula

$$V = \sum_{i=1}^n w_i s_i, \quad \sum_{i=1}^n w_i = 1 \quad (1)$$

where V denotes the overall viability score, s_i represents the normalized score of the i^{th} evaluation criterion, and w_i denotes its corresponding weight. These dimensions are as follows: market need, competitiveness, value proposition, technological feasibility, scalability, and business model suitability. This score is then used in conjunction with the individual dimension observations to formulate the overall viability evaluation.

3.5 Financial Assessment

The financial assessment module evaluates the set of financial metrics which can be obtained for the submitted startup. They include income, cash flow, burn rate, runway, cost per Customer Acquisition Cost (CAC), and the value of the customer LifeTime Value (LTV). These metrics are evaluated in combination with the viability assessment derived from the LLM. In case when both LTV and CAC are known, their ratio is defined by the following expression:

$$\frac{LTV}{CAC} = \frac{LTV}{CAC} \quad (2)$$

where LTV represents the estimated lifetime, value generated by a customer and CAC denotes the cost incurred to acquire that customer. The resultant indicator is part of the financial analysis process along with the cash flow position, burn rate, and runway. The process thus moves beyond assessing the viability of an idea to considering financial sustainability prior to moving on to risk assessment and recommendations.

3.6 Risk Analysis and Recommendation Generation

The risk analysis stage combines the outputs that have been derived from the multi-factor assessment of viability and financial analysis. This system analyzes different types of risks that may arise from the market, competitive situation, technical aspects, scalability, and financial

aspect considering the attributes of the startup and the observations by the LLM. Every single risk corresponds to the specific weakness of the startup idea analyzed.

From the analysis of the discovered weaknesses and the result of the assessment carried out, the LLM formulates recommendations that will be used in enhancing the concept of the startup. This is done by formulating recommendations based on the findings of the assessment process and not on a pre-established list of recommendations.

3.7 Report Generation, Storage, and Retrieval

Post viability, financial, and risk assessment of the system, the output is combined and compiled into a validation report. This report comprises the criterion level assessment, viability score, financial assessment, risks detected, and recommendations for improvement. The Flask backend sends the generated report back to the web interface where it can be presented to the user immediately.

At the same time, the final evaluation is stored in the MongoDB database, which ensures that past evaluations will be saved in the form of evaluation documents. Using the history interface, these saved evaluations will be fetched whenever needed so that users can retrieve their past evaluations without running the original idea again.

Algorithm 1: AI startup idea validation

Input: Startup idea X

Output: Structured validation report R

Step 1: Receive the startup inputs: problem statement, target market, proposed solution, and business model.

Step 2: Validate the submitted fields and construct the structured startup representation X_s .

Step 3: Prepare the structured prompt using X_s and the predefined evaluation criteria.

Step 4: Submit the prompt to the LLM through the Flask backend.

Step 5: Obtain criterion-level assessments for market demand, competition, value proposition, technical feasibility, scalability, and business-model suitability.

Step 6: Extract and normalize the generated criterion scores.

Step 7: Compute the overall viability score using Equation (1).

Step 8: Process the available financial parameters and calculate the LTV/CAC ratio using Equation (2) when both values are available.

Step 9: Identify market, competitive, technical, scalability, and financial risks from the assessment outputs.

Step 10: Generate recommendations corresponding to the identified weaknesses.

Step 11: Assemble the viability score, criterion assessments, financial analysis, risks, and recommendations into report R .

Step 12: Store R in MongoDB and display the completed report through the web interface.

END

4. Results and Discussion

4.1 Experimental Setup

The proposed system was tested by making use of the web-based AI Startup Idea Validator, which consists of a Flask backend, LLM API integration, MongoDB database storage, and finance evaluation module. Information related to startup ideas was provided through the web interface and then evaluated using the process outlined above in the methodology.

Table 2. System components and configuration

Component / Parameter	Configuration
Application Type	Web-based startup evaluation system
Backend Framework	Python with Flask
AI Processing	Large Language Model (LLM) API
Input Representation	Problem statement, target market, proposed solution, and business model
Evaluation Dimensions	Market demand, competition, competitive advantage, market opportunity, team capability, scalability, feasibility, and risk
Viability Assessment	Weighted multi-factor scoring

Financial Assessment	Revenue, profit margin, cash flow, burn rate, runway, funding, debt-to-equity, CAC, and LTV
Financial Indicator	<i>LTV / CAC</i> ratio
Database	MongoDB
Result Management	Persistent evaluation history and detailed-result retrieval
Benchmark Dataset	Kickstarter 2018 (Pre-Processed Data) [11]
Dataset Records	191,724 campaign records
Dataset Partition	80% development/reference, 20% held-out test
Partition Strategy	Stratified random split
Evaluation Mode	Real-time application execution and held-out benchmark assessment

4.2 LLM-Based Startup Evaluation Output

The experimental evaluation tests how the system converts a given startup idea into a multi-dimensional evaluation structure. For the edume startup idea, the report generated from the system gives a viability score of 56/100 and recommendation of MODERATE. The evaluation determines that EdTech is the targeted market and evaluates the idea using competitor analysis, competitive advantage, market opportunity, and team capability.

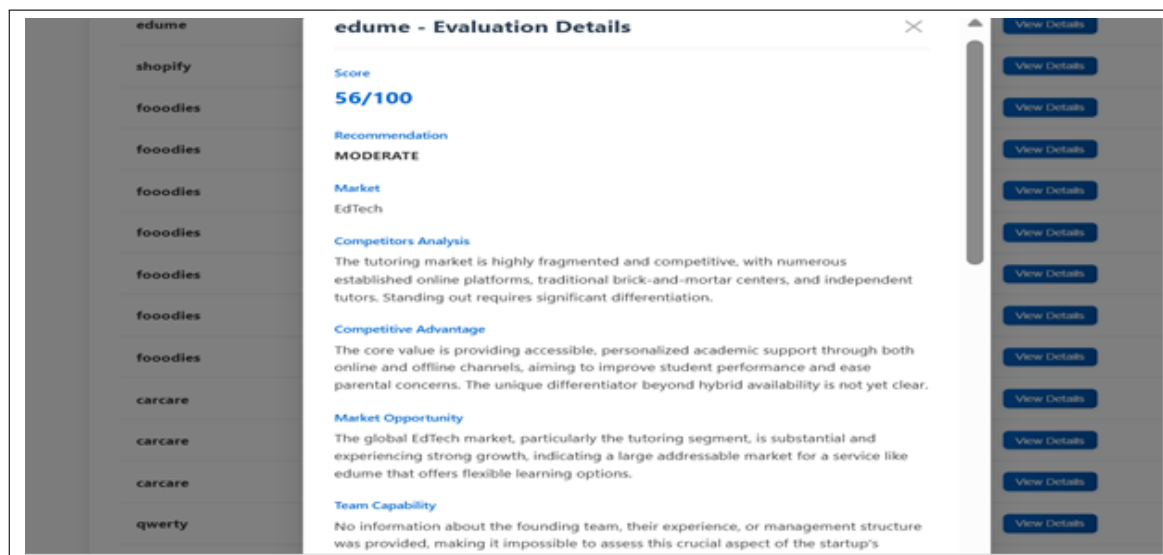


Figure 3. LLM-based startup evaluation output

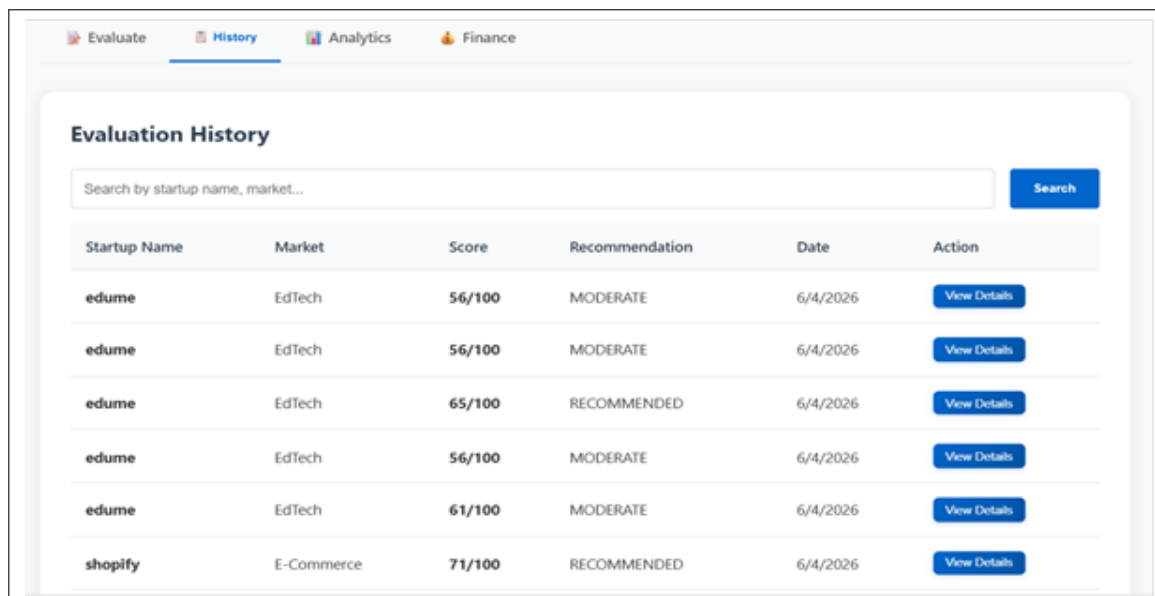
As shown in Figure 3. the competitor analysis reveals a fragmented market of tutoring,

where there are already online and offline options available, whereas the competitive advantage analysis reveals the constraint in the form of limited differentiation. On the other hand, the market opportunity analysis reveals a significant demand in the target market segment. The above example proves that the LLM module is not only providing one viability assessment but also breaking down the entire start-up idea into various criterion-based assessments.

4.3 Evaluation History and Persistent Result Management

The evaluation history module ensures a sustainable way of managing the startup evaluations created. All the evaluated startups are saved with their respective names, target markets, scores on viability, recommendations made, and dates on which the evaluations were processed.

As illustrated in Figure 4, the interface keeps several evaluation records, comprising evaluations of the edume startup done more than once and evaluation of Shopify. The keeping of each individual evaluation ensures that the results of the assessment can be compared across various submissions without having to undergo the whole LLM inference process again. The incorporation of MongoDB into the Flask back-end thus acts as the evaluation record management layer in the startup validation system.



The screenshot shows a web application interface with a navigation bar containing 'Evaluate', 'History', 'Analytics', and 'Finance'. The 'History' tab is active. Below the navigation bar is a section titled 'Evaluation History' with a search bar labeled 'Search by startup name, market...' and a 'Search' button. Below the search bar is a table with the following data:

Startup Name	Market	Score	Recommendation	Date	Action
edume	EdTech	56/100	MODERATE	6/4/2026	View Details
edume	EdTech	56/100	MODERATE	6/4/2026	View Details
edume	EdTech	65/100	RECOMMENDED	6/4/2026	View Details
edume	EdTech	56/100	MODERATE	6/4/2026	View Details
edume	EdTech	61/100	MODERATE	6/4/2026	View Details
shopify	E-Commerce	71/100	RECOMMENDED	6/4/2026	View Details

Figure 4. Persistent startup evaluation history

4.4 Financial Assessment and Sustainability Analysis

The financial analysis module analyzes the financial situation of the startup based on the parameters entered in the application. The financial analysis takes into consideration factors such as revenue generation, profit-making capacity, cash flow status, capital consumption, operational runway, funding model, and the economics of the customer.

As illustrated in Figure 5, edume is financially assessed to have moderate financial situation and sustainability. This can be attributed to the fact that in the analysis of the components, edume is found to have better unit economics and runway features, whereas profitability is found to be a weak aspect which needs to be improved. In this case, the system considers operational efficiency and cost optimization to be suitable solutions for the identified problems.

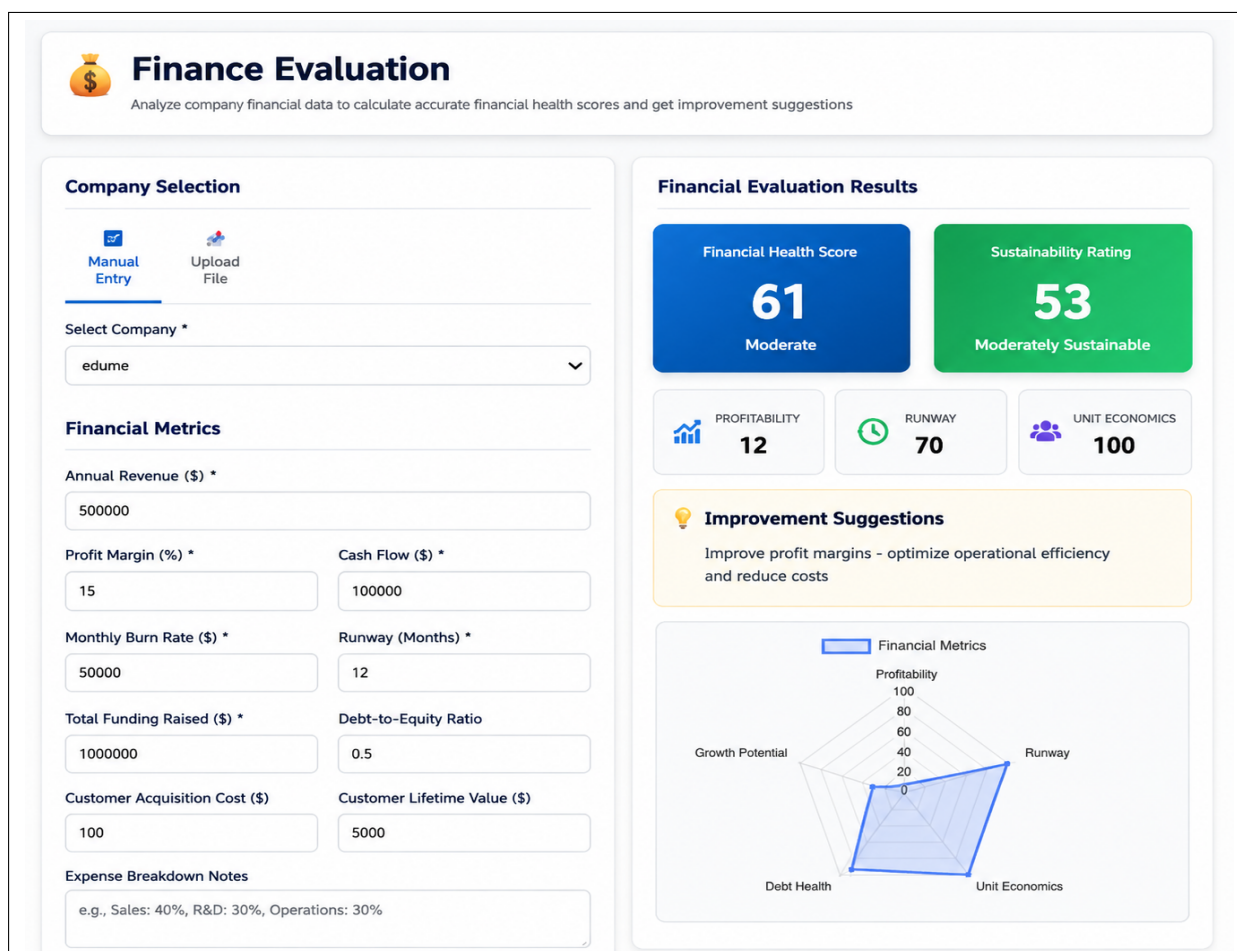


Figure 5. Startup financial sustainability assessment

4.5 Quantitative Evaluation Results

Quantitative assessment of the effectiveness of classification in the developed startup evaluation system is conducted using the comparison of the decision classes formed and the ground-truth decision classes described in section 3.2. This comparison considers three classes – Recommended, Moderate, and Not Recommended and analyzes how well the generated decision classes correspond to the reference decision classes. The class-wise distribution obtained in this way is shown in the form of the confusion matrix in Figure 6.

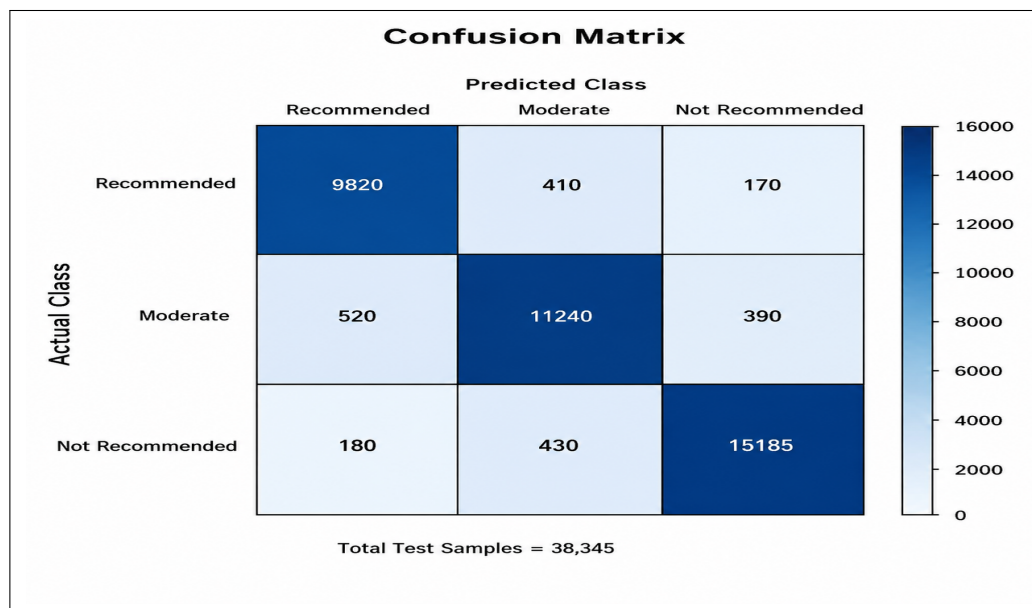


Figure 6. Confusion matrix of startup outcome assessment

The confusion matrix shows considerable clustering of the predictions on the main diagonal, implying consistency in agreement between the predictions made by the system and the corresponding references. The off-diagonal clustering is also modest, implying that the evaluation framework can effectively differentiate the key decision categories, although the overlapping near the middle category is due to the ambiguity of the borderline startups.

The derived performance measures are presented in Table 3. The high correlation between the macro-averaged metrics shows the consistent classification pattern within the three types of decisions. The most prominent class-level classification is achieved for the category "Not Recommended", while the lower value obtained for Moderate points at the fact that the difficult-to-categorize cases constitute the main challenge of the task. On the whole, the analysis demonstrates that the developed classification system allows the reliable separation of the classes.

Table 3. Classification performance by class

Class	Precision (%)	Recall (%)	F1-Score (%)
Recommended	93.35	94.42	93.88
Moderate	93.05	92.51	92.78
Not Recommended	96.44	96.14	96.29
Macro Average	94.28	94.36	94.32

5. Conclusion and Future Work

The proposed AI-powered startup validation system presents an example of an integrated method of evaluating startups at the initial stage by employing the following components: contextual reasoning with the help of LLM, viability assessment, financial assessment, and persistent management of evaluations. Specifically, the Flask framework is used for transforming the description of startups into structured data for analysis and providing insights on criteria such as demand, competitiveness, uniqueness, scalability, feasibility, and risks. The financial component of the system analyzes key economic factors and sustainability limitations. Such an integration allows for a realistic approach to assessing and improving startups prior to any considerable investment of resources. Future research may involve expansion of the system using model modification by domain, integration of real-time external market data, predicting outcomes, more sophisticated financial analysis, and empirical validation on a greater scale. Traceability and uncertainty estimation of explainable evaluations will increase their transparency and validity.

References

- [1] Kleinert, Simon, and Diemo Urbig. "Startup Evaluations with Generative Artificial Intelligence: An Exploratory Study on Early-Stage Investments and Survival Predictions by Large Language Models." *Entrepreneurship Theory and Practice*, 2026, 1-43.
- [2] Razaghzadeh Bidgoli, Mona, Iman Raeesi Vanani, and Mehdi Goodarzi. "Predicting the Success of Startups using a Machine Learning Approach." *Journal of Innovation and Entrepreneurship*, 20254, vol. 13, no. 1: 80.

- [3] Preuveneers, Jack, Joseph Ternasky, Fuat Alican, and Yigit Ihlamur. "Reasoning-based AI for Startup Evaluation (raise): A Memory-Augmented, Multi-step Decision Framework." ArXiv Preprint, 2025, arXiv:2504.12090.
- [4] Kim, Jongwoo, Hongil Kim, and Youngjung Geum. "How to Succeed in the Market? Predicting Startup Success using a Machine Learning Approach." *Technological Forecasting and Social Change*, 2023, vol. 193: 122614.
- [5] Li, Yisheng, Iman Zadehnoori, Ahmad Jowhar, Sean Wise, Andre Laplume, and Morteza Zihayat. "Learning from Yesterday: Predicting Early-Stage Startup Success for Accelerators through Content and Cohort Dynamics." *Journal of Business Venturing Insights*, 2024, vol. 22: e00490.
- [6] Wei, Chih-Ping, Evana Szu-Han Fang, Chin-Sheng Yang, and Pin-Jun Liu. "To Shine or not to Shine: Startup Success Prediction by Exploiting Technological and Venture-Capital-Related Features." *Information & Management*, 2025, vol. 62, no. 6: 104152.
- [7] Ahlgren, Thea Lovise, Helene Fønstelien Sunde, Kai-Kristian Kemell, and Anh Nguyen-Duc. "Assisting Early-Stage Software Startups with LLMs: Effective Prompt Engineering and System Instruction Design." *Information and Software Technology*, 2025, vol. 187: 107832.
- [8] Choi, Youngkeun. "Startup Success Prediction with PCA-Enhanced Machine Learning Models." *Journal of Technology Management & Innovation*, 2024, vol. 19, no. 4: 77-88.
- [9] Rady, Judy, David Townsend, and Rick Hunt. "From Algorithmic Hallucinations to Alien Minds: Addressing the Ideator's Dilemma through Entrepreneurial Work." *Journal of Business Venturing*, 2026, vol. 41, no. 1: 106550.
- [10] Pan, Chengen, Xiaohong Pan, and Li Sun. "A Multi-Layer AI Decision Support System for Startup Success Prediction and Risk Assessment using Knowledge Graphs and Federated Learning." *Scientific Reports*, 2026, vol. 16: 1-46.
- [11] H. Gunduz, "Kickstarter 2018 (Pre-Processed Data)," Zenodo, 2024.
<https://zenodo.org/records/12735215>