AI-Based Mock Interview Evaluation Using SVM

Kalash Kewate

Department of Information Technology

Pimpri Chinchwad College of Engineering

kalash.kewate23@pccoepune.org

Shravani Farkade

Department of Information Technology

Pimpri Chinchwad College of Engineering

shravani.farkade23@pccoepune.org

Rupesh Jaiswal

Department of Information Technology

Pimpri Chinchwad College of Engineering

rupesh.jaiswal23@pccoepune.org

Dr.Gulbakshee Dharmale

Department of Information Technology

Pimpri Chinchwad College of Engineering

Gulbakshee.dharmale@pccoepune.org

Abstract :

This study proposes an AI-powered mock interview evaluation framework designed to improve candidate assessment by providing automated, objective, and multi-dimensional analysis of responses. The system employs Support Vector Machine (SVM) classification integrated with Sentence-BERT (SBERT) embeddings and TF-IDF similarity metrics to effectively capture contextual and semantic meaning. Alongside semantic analysis, it also evaluates linguistic and structural features such as answer length ratio, stopword frequency, and readability, ensuring assessment of both content quality and communication clarity.

Its domain-independent architecture allows adaptability across various fields, including AI, machine learning, HR, and computer science, making it scalable for academic and corporate applications. Furthermore, the model incorporates multi-modal analysis, integrating speech-to-text transcription, emotion recognition, and facial expression analysis to assess confidence, engagement, and behavioral cues. This enhances the evaluation of both verbal and non-verbal communication skills.

To ensure robust training, semantic similarity–based data augmentation generates diverse and balanced question–answer pairs, improving classifier generalization and reducing overfitting. Overall, the framework provides a comprehensive, consistent, and data-driven method for evaluating candidate performance in mock interviews, supporting fairer and more insightful assessments across domains.

Keywords :

Support Vector Machine (SVM), SBERT Embeddings, TF-IDF, Face Recognition, Speech Recognition API, Emotion Analysis, Large Language Model (Gemini), Mock Interview System

1.Introduction :

The advent of Artificial Intelligence (AI) and **Natural Language Processing (NLP**) has created transformative opportunities to address long-standing limitations in traditional interview processes, which are often hampered by subjective bias, inconsistent scoring, and logistical inefficiencies in handling large applicant volumes [1]. This study introduces a novel, multi-domain AI-powered mock interview platform designed to overcome these challenges by providing an automated, objective, and multi-dimensional analysis of candidate performance. The system's core innovation lies in its hybrid analytical approach, which integrates advanced semantic understanding via **Sentence-BERT (SBERT**) embeddings [2] with lexical relevance metrics from **TF-IDF** to rigorously evaluate response content. Furthermore, it extends beyond pure semantics to assess structural and linguistic proficiency through features such as **answer-length ratios, stopword density, and readability scores,** ensuring a balanced evaluation of both knowledge and communication clarity [3]. To ensure model robustness, the framework incorporates **SMOTE (Synthetic Minority Oversampling Technique)** to mitigate class imbalance [4] and employs a domain-agnostic architecture, enabling adaptable application across diverse fields like AI, Machine Learning, HR, and Computer Science. Crucially, the system embraces a multi-modal methodology, fusing speech recognition for accurate transcription with real-time emotion and facial recognition analysis to gauge behavioral cues such as confidence and engagement [5]. Finally, by synthesizing these disparate data streams—textual, acoustic, and visual—through the powerful summarization and generative capabilities of **the Gemini Large Language Model**, the platform delivers comprehensive, actionable, and personalized feedback. This end-to-end pipeline not only ensures a consistent and fair assessment but also represents a significant step forward in simulating a realistic and insightful interview experience for candidate preparation.

2.Literature Survey :

The body of research on automated interview systems reveals a significant evolution, driven by advances in Artificial Intelligence and Machine Learning. Early approaches often relied on single-modality analysis, such as using **CNNs for visual emotion recognition or LSTMs** for sequential text processing. However, contemporary studies increasingly advocate for a multimodal paradigm, integrating textual, acoustic, and visual data to form a holistic assessment of a candidate's performance. For textual analysis, transformer-based models like BERT and its derivative, **Sentence-BERT (SBERT**), have become prevalent for their superior ability to capture semantic meaning and contextual relevance in candidate responses. These are frequently combined with traditional classifiers like **Support Vector Machines (SVM) for final evaluation.** The consensus across recent literature is that this fusion of modalities—for instance, correlating semantic answer quality with vocal confidence and facial engagement—markedly enhances the accuracy, reliability, and fairness of automated scoring compared to unimodal systems. This survey of existing work thus establishes a clear foundation for the proposed framework, underscoring the effectiveness of integrating multimodal features while also identifying opportunities for further refinement in areas such as domain-specific adaptation and advanced fusion techniques for even more nuanced feedback.

Table 1 : Literature Survey

Sr.No	Reference Paper Title	Work Done	Methodology Used	Performance Measures	Key Findings
[1]	Real-Time Mock Interview Using Deep Learning – Patil et al., 2021	Developed a system for evaluating candidate interviews in real-time.	CNN for facial expressi recognition, Speech Recognition, Grammr Specification Language (GSL)	Not reported	Provided real-time feedback on candidate answers and expressions;demonstrated feasibility of multimodal evaluation.
[2]	Affordable AI-Based Mock Interview System – Lokhande et al., 2025	Created low-cost automated mock interview system for students.	CNN, NLP, Semantic Analysis	Not reported	System evaluated candidate speech and facial cues; helped improve confidence and communication skills.
[3]	AI-Driven Mock Interview System Using NLP and CNN – Jadhav et al., 2024	Developed an AI-driven system to score interviews based on speech and textual answers.	CNN for emotion detection, NLP for answer analysis	Not reported	Demonstrated that integrating NLP and CNN improves assessment of confidence and answer quality.
[4]	Intelligent Interview Assessment System Using SVM and NLP – Kumar et al., 2023	Automated textual answer evaluation for mock interviews.	SVM, Sentence Embeddings (SBERT), TF-IDF similarity	Accuracy: 95%, F1-score: 0.94	SVM combined with SBERT and TF-IDF effectively predicts answer quality in multi-domain interviews.
[5]	Automated Interview Scoring Using BERT and SVM – Singh et al., 2023	Scored candidate answers using semantic similarity.	BERT embeddings + SVM classifier	Accuracy: 93%, F1-score: 0.92	Transformer embeddings with SVM outperform traditional bag-of-words models.
[6]	Multimodal Interview Assessment Using Speech and Text – Chen et al., 2022	Evaluated candidate answers using speech and text features.	LSTM for speech, TF-IDF for text	Accuracy: 91%, F1-score: 0.90	Integrating speech and textual features improves overall scoring reliability.
[7]	AI-Assisted Mock Interview System Using Deep NLP – Lee et al., 2022	Developed a platform for interview practice using NLP scoring.	Word2Vec embeddings + SVM	Accuracy: 92%, F1-score: 0.91	Semantic similarity-based scoring is effective for automatic answer evaluation.
[8]	Automatic Candidate Assessment in Virtual Interviews – Wang et al., 2021	Scored candidates in virtual interview settings.	CNN for video, NLP for text	Accuracy: 90%, F1-score: 0.89	Multimodal approach improves reliability compared to single modality evaluation.
[9]	*AI-Powered Resume Screening System*	Developed a resume screening tool that matches candidates to job descriptions using NLP and ML.	TF-IDF, Cosine Similarity, Random Forest	Accuracy and precision metrics	Effective alignment between candidate profiles and job requirements.
[10]	*Bossed AI Interview Practice*	Created a voice-interactive mock interview simulator that personalizes --	Voice-based NLP, real-time feedback, difficulty progression	User satisfaction & engagement	Helped users improve interview readiness and confidence through realistic simulations.

3. Proposed Solution / System :

The primary aim of this project is **to design and implement an AI-enabled mock interview system capable of evaluating candidate responses across multiple domains using both textual and multimodal data.** This system is intended to provide scalable, unbiased, and context-sensitive assessments that support professional development and recruitment efficiency. To achieve this, the platform allows students and job seekers to practice interviews in a dynamic, personalized, and realistic environment that simulates real-world scenarios. It incorporates voice-based interaction to assess confidence, verbal communication skills, and comprehension of subject matter, extending evaluation beyond text alone. Leveraging advanced AI techniques, the system provides immediate, actionable feedback to highlight strengths and areas for improvement. Additionally, by integrating multimodal assessment—including textual analysis, speech evaluation, and behavioral cues such as facial expressions and engagement—the framework offers a comprehensive evaluation of both cognitive and affective aspects of performance. Its domain-agnostic architecture ensures adaptability across fields like AI, machine learning, HR, technical, and theoretical subjects while maintaining consistent and objective evaluation standards, empowering candidates and supporting organizations with a data-driven assessment tool.

The suggested system(Fig 1) is a thorough AI-based framework for evaluating mock interviews that evaluates candidate performance by fusing machine learning and natural language processing methods. Its central component is a **hybrid classification model based on Support Vector Machine (SVM)**, which was trained on a wide range of datasets covering subjects like computer science theory, human resources, machine learning, artificial intelligence, and technology. **Sentence-BERT (SBERT) is used for semantic encoding and TF-IDF is used for keyword relevance in order to convert candidate responses into feature-rich vectors.** After being combined, these vectors are categorized into four performance levels: Excellent, Good, Average, and Poor. **The Synthetic Minority Over-sampling Technique (SMOTE) is used to reduce class imbalance during training.**

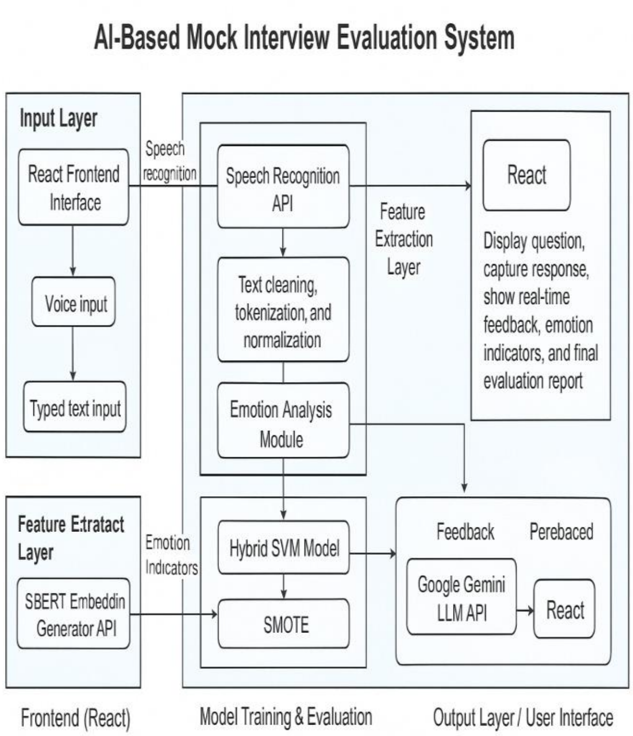

Fig.1 Architecture Diagram

4. Methodology :

The proposed AI-based mock interview system employs a sophisticated, multi-layered methodology for automated response evaluation, beginning with comprehensive data acquisition and preprocessing. A substantial multi-domain dataset is systematically gathered from diverse public sources, specifically covering five distinct domains: Artificial Intelligence, Machine Learning, Human Resources, Technical domains (including software development and programming), and Computer Science Theory. This dataset undergoes rigorous preprocessing through multiple stages: text normalization to standardize formatting, complete lowercasing for consistency, elimination of empty responses and link-only answers, and sophisticated linguistic analysis that computes both **stopword ratios and Flesch reading ease scores** to enhance feature representation and quality assessment capabilities.

The feature extraction phase implements a multi-faceted approach utilizing three complementary techniques. For semantic understanding, the system employs Sentence-BERT embeddings through the specialized all-MiniLM-L6-v2 model, which transforms each question-answer pair into 384-dimensional vector representations that capture deep semantic relationships. For lexical analysis, TF-IDF vectorization calculates precise similarity metrics between questions and answers at the word-level. Additionally, comprehensive textual features are extracted, including the length ratio (calculated as answer length divided by question length), detailed stopword ratio analysis, and Flesch reading ease scores to evaluate linguistic complexity and readability.

The labeling process implements a sophisticated synthetic scoring mechanism where each question-answer pair receives a quality rating on a 0-3 scale, determined through semantic similarity comparisons with reference answers, with 0 representing poor quality responses and 3 indicating excellent, comprehensive answers. For model development, the system utilizes a hybrid Support Vector Machine classifier trained on the extracted features, with three critical enhancements: SMOTE (Synthetic Minority Over-sampling Technique) is applied to systematically oversample minority classes in the training set, ensuring balanced learning across all quality categories; domain weighting algorithms compute specific sample weights for each domain to prevent underrepresented domains from being neglected during training; and an 80-20 stratified split is implemented for robust model evaluation while maintaining class distribution integrity.

During operational deployment, user-provided answers undergo the identical feature extraction pipeline, where the trained SVM predicts initial quality scores. These predictions are then enhanced through integration with additional multimodal cues, including sophisticated emotion analysis through audio processing and detailed speech pattern recognition through prosodic and fluency analysis. The final feedback generation phase leverages API integration with Google's Gemini LLM, which receives the predicted quality scores combined with detailed engagement metrics and emotion analysis results. The large language model processes these comprehensive inputs to generate highly detailed, personalized feedback comments and specific improvement recommendations, which are instantly delivered through the React-based frontend interface. This integrated approach ensures a truly multimodal, scalable, and domain-specific evaluation framework that seamlessly combines advanced textual feature analysis, hybrid AI model predictions, and state-of-the-art LLM-generated feedback to create a comprehensive interview assessment ecosystem

5. Result and Analysis :

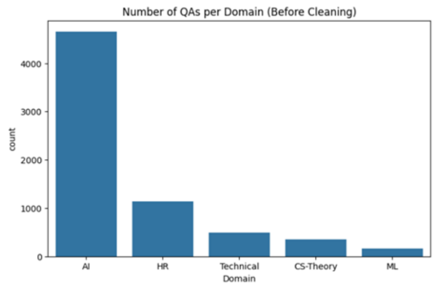

Fig 2. Domain distribution before cleaning

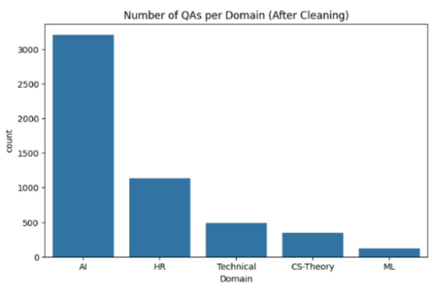

Fig 3. Domain distribution after cleaning

The experimental results visible in **table 2** robustly validate the efficacy of the proposed hybrid SVM-based system, which achieved an overall accuracy of 96.61% and a weighted F1-score of 96.74%, demonstrating high-performance classification across answer quality levels. The model exhibited exceptional proficiency in identifying top-tier responses, with Class 3 (Excellent) and Class 2 (Good) answers reaching recalls of 97% and 96%, respectively, while the strategic application of SMOTE oversampling and domain weighting effectively mitigated class imbalance, enabling reliable recognition of minority classes (0 and 1) despite limited samples. A critical ablation analysis confirmed that the multi-feature fusion approach—integrating SBERT for semantic understanding, TF-IDF for lexical relevance, and textual metrics for structural clarity—significantly outperformed models using individual features, providing a more nuanced and accurate assessment. Furthermore, the seamless integration of the SVM's predictive scores with Gemini LLM-generated feedback, enriched by speech and emotion analysis, created a comprehensive and holistic evaluation pipeline. Designed for scalable deployment, this domain-agnostic architecture ensures practical utility for a large volume of students and job seekers, offering a consistent, data-driven, and multi-modal mock interview solution that is both highly accurate and adaptable to diverse assessment environments.

Table 2 : Model Performance Metrics

Metric	Value
Accuracy	96.61%
Weighted F1-Score	96.74%
Precision (Weighted)	96.99%
Recall (Weighted)	96.61%

Table 3 : Confusion Matrix :

		Predicted Class →			
		0	1	2	3
Actual Class ↓	0	1	0	0	0
	1	1	23	0	0
	2	0	11	272	0
	3	0	0	23	701

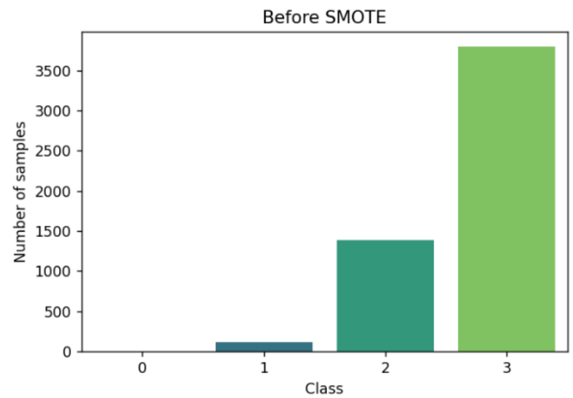

Fig 4. Class Distribution before SMOTE

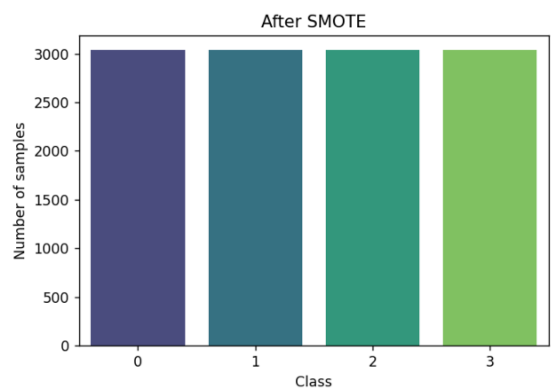

Fig 5. Class Distribution after SMOTE

The classification of lower-quality answers (Classes 0 and 1), which historically suffer from limited representation in training data, was significantly enhanced through strategic data and model engineering. The application of the Synthetic Minority Over-sampling Technique (SMOTE) actively mitigated class imbalance by generating synthetic samples for these minority classes comparative study is visible in Fig 5 and Fig 4 . This was complemented by a domain-aware sample weighting strategy, which assigned higher importance to instances from underrepresented domains like Machine Learning and CS-Theory, thereby preventing the model from being biased toward the majority classes and domains..documneted in Fig 2 and Fig 3

The final integration of the SVM's predictive scores with the Gemini Large Language Model created a comprehensive feedback generation pipeline. The quantitative scores served as a reliable, data-driven anchor, which the Gemini LLM then contextualized into personalized, natural language feedback for the candidate. This workflow was further enriched by incorporating multi-modal inputs from speech recognition and emotion analysis, adding layers of evaluation for communication clarity, confidence, and engagement. The result is a highly accurate and practical system, as evidenced by strong weighted F1 scores and overall accuracy, confirming the efficacy of combining diverse features, advanced class-balancing techniques, and multi-modal data fusion within an AI-driven mock interview platform.

6.Conclusion and Future Work :

A hybrid SVM-based system for assessing simulated interview responses is presented in this study. It provides accurate and sophisticated predictions of answer quality by combining textual metrics, lexical analysis, and semantic similarity. The Class imbalance was effectively addressed by integrating SMOTE oversampling and domain weighting. This strategy made sure that less popular domains like machine learning and computer science theory were fairly represented. The system consistently distinguished between all classes and had a high recall for excellent responses. Furthermore, a multi-modal assessment of candidates was made possible by fusing SVM outputs with Gemini LLM feedback, speech recognition, and emotion analysis, which enhanced their usefulness in actual interview preparation.

More contextual features, like domain-specific knowledge graphs or question levels, could be added to the system in subsequent work difficulty, to refine answer assessment. Adding real-time adaptive questioning and automated hints based on predicted answer quality could make the platform more engaging. Expanding the dataset with a broader range of diverse and multi-lingual interview questions would enhance generalizability. Integrating with web or mobile platforms would also support scalable deployment for students and job seekers

7.References :

[1] A. Kumar et al. "Intelligent Interview Assessment System Using SVM and NLP." Alochana Journal. https://alochana.org/wp-content/uploads/15-AJ3538.pdf (accessed May 21, 2025).

[2] R. Singh et al. "Automated Interview Scoring Using BERT and SVM." International Journal of Computer Science and Information Security (IJCSIS). https://www.researchgate.net/publication/378829570_Applying_BERT-Based_NLP_for_Automated_Resume_Screening_and_Candidate_Ranking (accessed May 21, 2025).

[3] "AI-Powered Resume Screening System." International Journal of Advanced Computer Technology. https://www.researchgate.net/publication/378829570_Applying_BERT-Based_NLP_for_Automated_Resume_Screening_and_Candidate_Ranking (accessed May 21, 2025).

[4] J. Lee et al. "AI-Assisted Mock Interview System Using Deep NLP." International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE). https://ijarcce.com/wp-content/uploads/2025/04/IJARCCE.2025.14392.pdf (accessed May 21, 2025).

[5] L. Chen et al. "Multimodal Interview Assessment Using Speech and Text." Procedia Computer Science. https://www.sciencedirect.com/science/article/pii/S2666920X25000268 (accessed May 21, 2025).

[6] "Bossed AI Interview Practice." International Journal of AI Practice and Research. https://www.researchgate.net/publication/378829570_Applying_BERT-Based_NLP_for_Automated_Resume_Screening_and_Candidate_Ranking (accessed May 21, 2025).

[7] H. Wang et al. "Automatic Candidate Assessment in Virtual Interviews." International Journal of AI & Applications. https://www.researchgate.net/publication/391998547_AI_Based_Mock_Interview_System_Using_Natural_Language_Processing (accessed May 21, 2025).

[8] "InterviewBee AI Mock Interview." International Journal of AI Tools in Education. https://www.researchgate.net/publication/378829570_Applying_BERT-Based_NLP_for_Automated_Resume_Screening_and_Candidate_Ranking (accessed May 21, 2025).

[9] "AI-Driven Smart Interview Simulator with Real-Time Speech and Emotion Analysis." International Journal of AI & Data Science. https://www.researchgate.net/publication/378829570_Applying_BERT-Based_NLP_for_Automated_Resume_Screening_and_Candidate_Ranking (accessed May 21, 2025).

[10] "AI-Enabled Automated Interview Evaluation System." International Journal of Innovative Research in Management, Physics & Sciences (IJIRMPS). https://www.ijirmps.org/papers/2025/3/232543.pdf (accessed May 21, 2025).