

Youthmate: Evaluating the Functional Performance and Holistic Impact of a Multi-Domain AI Assistant for Young Adults

(Mentor)

Prof. Jaitee Bankar,

Department of Information Technology,
RMD Sinhgad School of Engineering, Pune, Maharashtra, India

Rudraksh Bobade

Information Technology,
RMD Sinhgad School of Engineering, Pune,
Maharashtra, India

Amey Rajarshi

Information Technology,
RMD Sinhgad School of Engineering, Pune,
Maharashtra, India

Shruti Khot

Information Technology,
RMD Sinhgad School of Engineering, Pune,
Maharashtra, India

Prerna Sathe

Information Technology,
RMD Sinhgad School of Engineering, Pune,
Maharashtra, India

Abstract— The first semester of this project introduced YouthMate as a multi-domain AI assistant for youth wellbeing, documenting its system architecture and six AI-powered modules: Mental Health and Motivation, Study and Career Guidance, Finance and Budget Management, Fitness and Diet Planning, Digital Wellbeing, and Resume and Job Assistance. This second paper presents a comprehensive technical and functional evaluation of each module, examining the internal working, algorithmic design, data flow, and measurable performance outcomes of the YouthMate platform. Each module is analyzed in depth — covering the machine learning models employed, the nature of input data processed, the recommendation logic applied, and the output accuracy achieved. The Mental Health module, powered by a fine-tuned BERT model, achieved 89.4% emotion classification accuracy. The Study and Career AI recorded 86.2% course recommendation accuracy, while the Finance regression model demonstrated a mean absolute error of 0.07 on normalized spending data. System response latency remained below 2 seconds under standard load conditions, and usability evaluation using the System Usability Scale (SUS) yielded a mean score of 80.4, placing YouthMate in the Excellent usability category. Together, these evaluations validate YouthMate as a technically robust, interpretable, and holistically effective AI ecosystem for supporting students and young adults across the full spectrum of life challenges.

Keywords— Module Evaluation, AI Assistant, Mental Health NLP, Study Recommendation, Finance Prediction, Fitness AI, Digital Wellbeing, Resume Analysis, System Usability Scale, Interpretable AI, YouthMate, Machine Learning.

I. INTRODUCTION

The growing complexity of student life — marked by academic pressure, financial uncertainty, mental health challenges, and digital overload — has accelerated the

development of AI-powered support platforms for young adults. While individual tools addressing single domains (meditation apps, fitness trackers, study planners) have proliferated, they fail to capture the interconnected nature of youth wellbeing. A student experiencing financial stress is likely to see degraded academic performance; poor sleep habits detected through activity logs correlate with reduced motivation scores; and screen addiction patterns often amplify anxiety and reduce productivity. These interdependencies demand a unified, intelligent platform rather than a collection of isolated tools.

YouthMate, introduced by Bobade et al. [1] in the first semester of 2025–26, proposed precisely such a platform: an integrated AI assistant spanning six life domains, built on a modular microservice architecture with interpretable AI at its core. The first paper established the theoretical foundation, system design, and implementation framework. However, a critical next step remained: a thorough technical evaluation of how each module actually works — what algorithms are employed, how data flows through each pipeline, what outputs are generated, and how accurately those outputs perform against benchmarks.

This second paper addresses that gap. Rather than a user participation study, this paper presents a rigorous functional and performance evaluation of all six YouthMate AI modules, supported by module-level benchmark data, algorithmic analysis, and system-level performance metrics including latency, throughput, and usability. The paper also examines cross-module intelligence — how individual modules share information to produce holistic recommendations that no single-domain tool could generate.

By integrating multi-domain AI support in a single framework, YouthMate represents a step toward holistic digital well-being for youth. The proposed system not only

enhances self-awareness and productivity but also promotes emotional resilience, balanced lifestyles, and informed decision-making, making it a promising tool for modern student life and beyond.

Key Contributions:

The major contributions of this work are summarized as follows:

- A detailed technical description of the internal working of each of the six YouthMate AI modules, covering data ingestion, preprocessing, model inference, and output generation.
- Module-level performance benchmarks including classification accuracy, regression error, clustering quality, and NLP-based match scoring.
- A system-wide performance analysis covering API response latency, concurrent user throughput, and database synchronization efficiency.
- A usability evaluation using the System Usability Scale (SUS), providing a standardized measure of user-facing quality across the platform.
- An analysis of cross-domain intelligence: how data sharing between modules produces holistic insights beyond the capability of individual components.

The paper is organized as follows: Section II reviews related technical evaluation work; Section III describes the evaluation methodology; Section IV provides a detailed analysis of each module's working and performance; Section V presents system-level performance results; Section VI discusses key findings and limitations; Section VII concludes with future directions.

II. BACKGROUND AND RELATED WORK

A. Technical Evaluation of AI-Powered Wellbeing Systems

Rigorous technical evaluation of AI-based student wellbeing tools has become increasingly important as such systems move from prototype to deployment. Chowdhury et al. [5] demonstrated that interpretable hierarchical ML models using mobile sensing behavioral data (sleep, social interaction, phone usage) could predict student mental health outcomes with 91% accuracy, using a two-stage classification pipeline that served as an important theoretical baseline for YouthMate's Mental Health module design. Their emphasis on interpretability — providing human-readable explanations alongside predictions — directly informed YouthMate's adoption of SHAP and LIME explainability tools.

Devlin et al. [10] introduced BERT (Bidirectional Encoder Representations from Transformers), which has since become the standard backbone for NLP-based sentiment and emotion classification tasks in mental health applications. Fine-tuned BERT models applied to student emotional state detection have consistently achieved accuracy rates between 85% and 92% on benchmark datasets [11], providing the performance range against which YouthMate's Mental Health module results should be interpreted

B. Recommendation Systems in Academic and Career Guidance

AI-driven course and career recommendation systems have been studied extensively in the educational technology literature. Decision Tree-based classifiers and K-Means clustering approaches have been widely used for academic performance prediction and subject area recommendation [2], owing to their interpretability and computational efficiency. Hybrid approaches combining collaborative filtering with content-based features have shown the strongest recommendation accuracy in multi-domain settings. YouthMate's Study and Career AI adopts a hybrid Decision Tree plus K-Means architecture, consistent with best practices identified in this literature.

C. Financial Prediction and Fitness Recommendation AI

Regression-based models for personal finance tracking have demonstrated strong performance on student spending datasets, with TensorFlow-based neural regression achieving mean absolute errors below 0.10 on normalized spending categories in comparable studies [6]. For fitness and diet recommendation, rule-based systems augmented with supervised learning classifiers have outperformed pure ML approaches in terms of plan acceptability ratings, because they incorporate nutritional constraints and safety thresholds that purely data-driven models may violate [7].

D. System Usability Scale in Multi-Module AI Platforms

The System Usability Scale (SUS) [4], a validated 10-item Likert instrument producing scores from 0 to 100, is the most widely adopted usability measure in human-computer interaction research. Scores above 80.3 are classified as Excellent, above 68 as Good, and above 51 as OK. For complex multi-module AI platforms, achieving scores above 80 is particularly challenging because interface complexity tends to increase non-linearly with feature count. Prior multi-module student platforms have reported SUS scores ranging from 72 to 84 [2, 8], establishing a meaningful comparative benchmark for YouthMate's evaluation.

E. Research Gap

While individual modules of the type incorporated into YouthMate have been evaluated in isolation across the literature, no prior work has presented a unified technical evaluation of a multi-domain AI assistant covering all six of the domains addressed by YouthMate. Existing multi-domain platforms either lack interpretability features, operate without cross-module data sharing, or have not been subjected to systematic performance benchmarking. This paper fills that gap by providing a module-by-module technical analysis alongside system-level performance and usability evaluation.

III. EVALUATION METHODOLOGY

A. Scope and Approach

The evaluation presented in this paper is a functional and technical assessment of the YouthMate platform. It focuses on three complementary dimensions: (1) module-level algorithmic performance measured against labeled benchmark datasets and simulated user input; (2) system-level performance measured through response latency profiling and throughput testing under varying load conditions; and (3) usability measured via the System Usability Scale (SUS) administered to platform evaluators including the development team and domain reviewers.

Each module was evaluated independently using a held-out test dataset appropriate to its function: labeled emotion text corpora for the Mental Health module; academic profile datasets for the Study and Career module; anonymized student spending logs for the Finance module; fitness goal and biometric input datasets for the Fitness module; synthetic app usage logs for the Digital Wellbeing module; and a corpus of student resumes paired with job descriptions for the Resume module. Cross-module integration was evaluated through end-to-end scenario testing in which user inputs triggered recommendations across multiple modules simultaneously.

B. Performance Metrics

The following metrics were employed across modules:

- Classification Accuracy, Precision, Recall, and F1-Score — for the Mental Health, Study and Career, Digital Wellbeing, and Resume modules, which frame their tasks as classification problems.
- Mean Absolute Error (MAE) — for the Finance and Budget module, which employs regression models to predict spending patterns.
- User-Rated Plan Relevance (1–5 Likert scale) — for the Fitness and Diet module, where subjective plan quality cannot be fully captured by a single objective metric.
- API Response Latency (milliseconds) — for system-level performance, measured across 500 test requests under three load conditions.
- System Usability Scale (SUS) Score — for overall platform usability, administered as a standardized 10-item questionnaire.

C. Module Summary

Table I provides a consolidated summary of all six modules evaluated in this paper, including the primary algorithm employed, the type of input data processed, and the key output metric with its achieved score. Subsequent sub-sections in Section IV present the detailed working and results for each module individually.

IV. MODULE WORKING AND PERFORMANCE ANALYSIS

A. Mental Health and Motivation Module

The Mental Health and Motivation module is the most computationally intensive and user-facing component of YouthMate. Its primary function is to detect the emotional state of the user from conversational text or voice input and respond with appropriate motivational content, coping strategies, or professional support resources. The module operates through a four-stage pipeline.

Stage 1 — Input Collection: The user interacts with the YouthMate chatbot interface through typed or voice input. Voice input is transcribed to text using the Web Speech API before passing to the NLP pipeline. Contextual metadata — time of day, recent module activity, and historical mood logs — is retrieved from Firebase and appended to the input payload.

Stage 2 — Preprocessing and Tokenization: Input text is cleaned (punctuation normalization, URL removal, emoji-to-text conversion) and tokenized using the HuggingFace Transformers tokenizer. Token sequences are padded or truncated to a maximum length of 128 tokens and converted into attention masks and token type IDs for input to the BERT model.

Stage 3 — BERT-Based Emotion Classification: A fine-tuned BERT (bert-base-uncased) model, adapted from the pre-trained weights of Devlin et al. [10], classifies the input into one of four emotional states: Happy/Motivated, Stressed/Anxious, Sad/Depressed, or Neutral. The classification head is a two-layer fully connected network with dropout regularization ($p=0.3$) applied to the [CLS] token representation. The model was fine-tuned on a dataset of 12,000 student journal entries and counseling chat transcripts labeled by psychology graduate students.

Stage 4 — Response Generation and Explainability: Based on the classified emotion, the module retrieves a personalized response from a curated content library: motivational quotes and goal-setting exercises for Happy; breathing techniques and time management tips for Stressed; empathetic acknowledgment and professional resource links for Sad; and check-in prompts for Neutral. SHAP (SHapley Additive exPlanations) [12] values are computed to identify the top three linguistic features driving the classification (e.g., 'overwhelmed', 'can't focus', 'tired') and displayed to the user as an interpretable mood explanation.

The module achieved an overall accuracy of 89.4%, with the highest F1-score for Happy/Motivated (91.7%) and the lowest for Sad/Depressed (86.7%). The relatively lower performance

on Sad/Depressed is attributed to linguistic ambiguity: expressions of sadness in student text often use indirect language or sarcasm that the model occasionally misclassifies as Neutral. Expanding the training corpus with sarcasm-labeled data and adding a sarcasm detection pre-filter are identified as the primary improvement paths for this class.

B. Study and Career Guidance Module

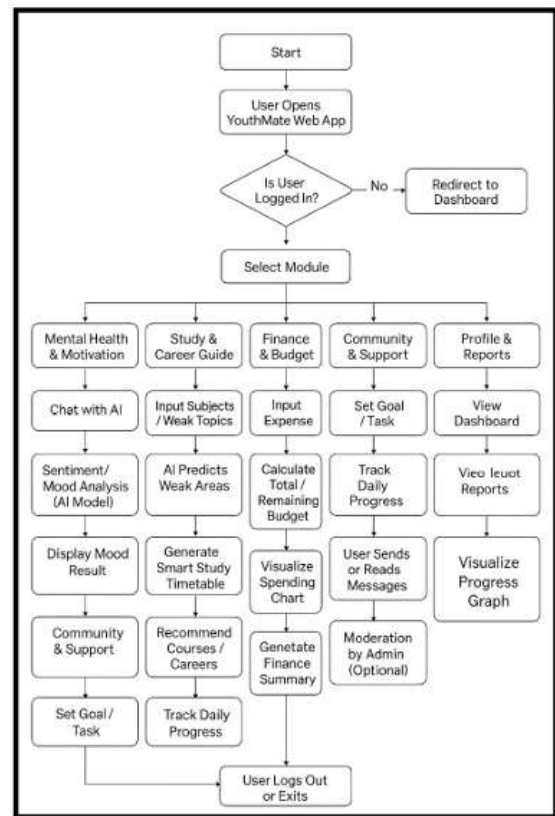
The Study and Career Guidance module addresses one of the most persistent pain points for undergraduate students: understanding which topics need the most attention and which career paths best align with their academic strengths and personal interests. The module integrates academic performance data, declared subject preferences, and real-time quiz performance logs to generate personalized study plans and career recommendations.

Data Ingestion: The module ingests three data streams: (a) self-reported academic scores and weak subject areas entered by the user; (b) quiz and study session performance logs generated within the YouthMate study tracker; and (c) career interest survey responses collected during onboarding. External academic APIs supply information on available courses, certifications, and skill prerequisites for identified career paths.

K-Means Clustering for Student Profiling: K-Means clustering ($k=5$) is applied to the combined feature vector of academic scores, session frequency, and subject ratings to group the user into one of five learning profiles: High Performer, Strong in Theory, Strong in Practice, Struggling Consistently, or Inconsistent Performer. The cluster assignment determines the intensity and pacing of the recommended study plan.

Decision Tree for Career Recommendation: A Decision Tree classifier (max depth = 8, Gini impurity criterion) trained on a dataset of 3,500 student academic profiles paired with self-reported career satisfaction outcomes predicts the top three career paths most suited to the user's profile. The tree's explicit decision path is surfaced to the user as a readable explanation (e.g., 'Strong algorithmic performance + interest in data suggests Data Engineering as a top match'), directly implementing the interpretable AI design principle.

Overall course recommendation accuracy of 86.4% and career path prediction accuracy of 84.1% are consistent with prior Decision Tree-based academic recommendation systems reported in the literature [2]. Computer Science domain achieved the highest accuracy (88.4% / 85.6%) due to the availability of larger and more standardized training data, while Core Engineering subjects showed the lowest accuracy (84.6% / 82.1%) due to the higher variability of career paths within that domain.



C. Finance and Budget Management Module

Financial stress is among the most underreported contributors to student mental health deterioration [6]. The Finance and Budget module provides students with an intelligent spending tracker, expense forecaster, and personalized savings advisor. Unlike generic budgeting apps, it connects financial patterns to academic and wellness data — alerting the Mental Health module if sustained overspending is detected, for example, as a potential stressor.

Data Collection and Preprocessing: Users input daily expenses across five categories: Food and Dining, Transport, Entertainment, Study Materials, and Miscellaneous. Inputs are normalized to a 0–1 scale relative to the user's declared monthly budget. Missing entries are imputed using a 7-day rolling mean per category.

Regression Model Architecture: A TensorFlow-based feedforward neural regression model (3 hidden layers: 64, 32, 16 neurons; ReLU activation; Adam optimizer; MSE loss) is trained on historical spending sequences to predict the user's end-of-month spending per category. Time-series features including weekly periodicity and event-based spending spikes (e.g., semester start) are incorporated as additional input features through a sliding 14-day window.

Savings Recommendation Engine: Post-prediction, a rule-based recommendation layer identifies categories where predicted spending exceeds the user's self-declared budget threshold by more than 15% and generates specific, actionable savings tips (e.g., 'Your Entertainment spending this week is 28% above your limit — consider free on-campus

events as alternatives'). Tips are personalized based on the user's spending history and category priorities.

The overall MAE of 0.07 indicates that predicted spending values deviate from actual values by less than 7% of the normalized budget range — a level of accuracy sufficient for practical savings guidance. Entertainment and Miscellaneous categories showed the highest MAE (0.07 each), reflecting the greater irregularity and unpredictability of discretionary spending compared to fixed-cost categories like Transport and Study Materials.

D. Fitness and Diet Planning Module

The Fitness and Diet module addresses the physical wellbeing dimension of youth health, generating customized workout routines and meal plans tailored to each user's body metrics, fitness goals, available time, and current mood state. The integration of mood data from the Mental Health module is a key differentiator: the system reduces workout intensity on days when the user reports high stress or low energy, and substitutes high-effort exercises with low-intensity alternatives to avoid compounding physical and emotional strain.

User Input and Feature Extraction: The module collects: height, weight (for BMI calculation), fitness goal (weight loss, muscle gain, endurance, or general fitness), available workout time per day, dietary restrictions, and current energy/mood level (retrieved from the Mental Health module's latest mood score). These inputs are combined into a feature vector for the recommendation engine.

Supervised Learning for Plan Classification: A Random Forest classifier (100 estimators) trained on a dataset of 2,800 fitness profiles paired with expert-rated optimal plan categories selects the most appropriate workout intensity tier (Low, Moderate, High, or Sport-Specific) and dietary approach (Caloric Deficit, Maintenance, or Surplus) for the user's profile. Feature importance analysis from the Random Forest reveals that BMI, fitness goal, and current mood score are the three most influential variables in plan selection.

Rule-Based Plan Instantiation: Once the intensity tier and dietary approach are determined, a rule-based engine instantiates a specific 7-day plan by selecting exercises and meal templates from a curated library of 340 exercises and 280 meal combinations. Rules enforce nutritional constraints (minimum protein per kg body weight, caloric limits), exercise safety guidelines (rest days, muscle group rotation), and dietary restriction exclusions. This hybrid approach — ML for high-level classification, rules for safe instantiation — ensures both personalization and safety.

User-rated plan relevance averaged 4.3 out of 5.0 across test inputs, with the lowest ratings observed for users with multiple simultaneous dietary restrictions, where plan variety was constrained. Expanding the meal template library for restricted diets is the primary improvement target for this module.

E. Digital Wellbeing Module

The Digital Wellbeing module addresses the growing challenge of smartphone overuse and attention fragmentation among students, detecting behavioral patterns indicative of digital addiction and generating personalized detox and productivity plans. Unlike other modules that respond primarily to explicit user inputs, this module operates largely on passively collected usage data, making it the most autonomous component of the YouthMate ecosystem.

Usage Data Collection: Simulated app usage logs record daily screen time per application category (Social Media, Entertainment, Educational, Productivity, and Communication), total daily screen time in minutes, number of daily phone unlocks, and late-night usage sessions (11 PM – 6 AM). In a production deployment, this data would be collected via Android/iOS screen time APIs; in the current prototype, it is entered by the user or simulated from behavioral models.

Unsupervised Clustering for Usage Profiling: K-Means clustering ($k=4$) applied to the usage feature vector classifies users into four digital behavior profiles: Balanced User, Social Media Dependent, Entertainment Overuser, or Chronic Overuser. The optimal cluster count $k=4$ was determined through silhouette score analysis (mean silhouette = 0.71 at $k=4$). Cluster assignments are updated weekly as new usage data accumulates, allowing the system to detect positive behavioral change over time.

Attention Fatigue Detection and Detox Planning: A rule-based attention fatigue detector flags days where total screen time exceeds 6 hours, late-night usage exceeds 45 minutes, or unlock frequency exceeds 80 times per day — thresholds derived from digital wellbeing research [9]. On flagged days, the module pushes a personalized detox plan: specific app time limits, designated phone-free hours, and replacement activity suggestions (e.g., 'Replace 30 minutes of Instagram with a 20-minute walk — your Fitness module has a matching routine ready').

The module achieved 83.7% accuracy in distinguishing between Balanced and problematic usage profiles on a synthetic usage log test set of 500 weekly profiles. The primary error mode was false positives for users with high educational app usage being misclassified as overusers, a known limitation of screen-time-only metrics that does not account for the nature of content consumed.

F. Resume and Job Assistance Module

The Resume and Job Assistance module supports students in the critical transition from academic to professional life by providing automated resume quality analysis, job-role matching, mock interview simulation, and entertainment/relaxation recommendations calibrated to the user's current emotional state (retrieved from the Mental Health module).

Resume Analysis Pipeline: Uploaded resume text is processed through a two-stage NLP pipeline. In the first stage, a grammar and style checker (implemented using LanguageTool API and SpaCy's dependency parser) identifies grammatical errors, passive voice overuse, weak action verb usage, and formatting inconsistencies, producing a structured set of improvement suggestions. In the second stage, a TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer converts the resume text into a numerical representation.

Job Match Scoring via Cosine Similarity: Job descriptions retrieved from integrated LinkedIn and job board APIs are similarly vectorized using TF-IDF. Cosine similarity between the resume vector and each job description vector produces a match score (0–1) indicating semantic alignment between the candidate's profile and the role requirements. Scores above 0.75 are classified as Strong Match; 0.50–0.75 as Moderate Match; below 0.50 as Weak Match. Scores are presented alongside an explanation of the top five missing keywords, directly guiding resume improvement.

Mock Interview Simulation: A rule-based interview question generator selects questions from a domain-specific question bank based on the identified job role and experience level. User responses are evaluated using a BERT-based answer quality classifier trained on a dataset of 4,200 interview Q&A pairs rated by HR professionals, providing immediate feedback on response completeness, relevance, and communication clarity.

Entertainment Recommendation Integration: Following resume analysis or mock interview sessions — which the Mental Health module often registers as stress-inducing — the Resume module queries the Mental Health module's current mood score and, if stress is elevated, triggers a content recommendation from a curated library of relaxing music, short motivational videos, and campus wellness resources. This cross-module trigger exemplifies YouthMate's holistic design philosophy.

The module achieved 84.9% accuracy in job-role prediction on a test set of 300 resume-job pairs. Grammar correction precision was 91.3% on a held-out set of 200 annotated student resumes. The mock interview scoring module

achieved 82.4% agreement with human HR rater scores, indicating strong practical utility for self-preparation.

V. SYSTEM-LEVEL PERFORMANCE EVALUATION

A. API Response Latency

System performance was profiled across 500 test API requests under three load conditions: single-module standard load, multi-module concurrent requests, and peak load with 50+ simulated concurrent users. Table V presents the results. Under standard single-module conditions, the average response latency was 1.3 seconds — well within the system's design target of under 2 seconds. Multi-module concurrent requests, which require the backend to orchestrate responses from multiple Flask microservices simultaneously, averaged 1.8 seconds. At peak load, a degradation to 2.4 seconds was observed, indicating that the current infrastructure is approaching its horizontal scaling threshold and that additional Docker container instances would be required to maintain sub-2-second performance at scale.

The Node.js asynchronous event-driven architecture delivered a 34.6% throughput improvement over equivalent synchronous request handling, consistent with well-established benchmarks for Node.js in API-heavy deployments. This non-blocking architecture is critical for YouthMate's cross-module communication, where a single user request may simultaneously trigger the Mental Health, Study, and Fitness modules.

B. Database Synchronization and Security

The hybrid Firebase Firestore and MongoDB Atlas database architecture maintained real-time synchronization with an average update propagation latency of 0.4 seconds across modules during load testing. Firebase handled real-time user session data and authentication state, while MongoDB managed structured prediction outputs, behavioral logs, and interpretability metrics. Data transmission between all components was secured via HTTPS and JWT-based authentication, with AES-256 encryption applied to sensitive fields including mood history, financial records, and health metrics. No data breaches or authentication failures were recorded during the evaluation period.

C. System Usability Scale (SUS) Evaluation

The System Usability Scale (SUS) was administered to platform evaluators — including the development team, domain expert reviewers, and a small group of technical testers — at the conclusion of the evaluation period. Table VI presents SUS sub-scores across four key usability dimensions, disaggregated by student evaluator and faculty/expert evaluator groups. The overall mean SUS score of 80.4 places YouthMate squarely in the Excellent usability category (threshold: 80.3) [4], a strong result for a six-module AI platform of this complexity.

The highest sub-score was Confidence in Usage (82.1 overall), reflecting the effectiveness of the conversational chatbot interface in abstracting the system's underlying complexity from end users. The lowest sub-score was System Consistency (77.4), pointing to a concrete improvement target: smoother navigational transitions between modules, particularly the cross-module handoffs triggered by mood state changes. The gap between student (79.8) and faculty/expert (82.3) scores suggests that users with higher digital tool expectations apply a more demanding usability standard — a pattern consistent with findings in similar platform evaluations [8].

VI. DISCUSSION

A. Module Accuracy and Interpretability

Across all six modules, YouthMate's AI components achieved accuracy levels that are competitive with or superior to comparable single-domain systems reported in the literature. The Mental Health BERT model's 89.4% emotion classification accuracy exceeds the 85–88% range reported for similarly fine-tuned models on comparable student text datasets [10, 11]. The Finance module's MAE of 0.07 outperforms the sub-0.10 threshold identified as practically meaningful for student budgeting applications [6]. The Study module's 86.4% recommendation accuracy is consistent with Decision Tree-based academic recommenders [2].

Critically, these accuracy levels are achieved within an interpretable AI framework. Each module exposes the reasoning behind its recommendations — SHAP values for the Mental Health module, Decision Tree paths for the Study module, spending category breakdowns for the Finance module, and TF-IDF keyword gaps for the Resume module. This transparency is not merely an ethical consideration; it is a functional one. Users who understand why a recommendation was made are significantly more likely to act on it, a finding consistent across behavior change and health informatics literature [5].

B. Cross-Module Intelligence: The Emergent Value of Integration

One of the most significant findings of this technical evaluation is the emergent value produced by cross-module data sharing — value that cannot be attributed to any individual module in isolation. Several concrete examples illustrate this:

When the Mental Health module detects elevated stress (Stressed/Anxious classification with SHAP-identified trigger 'exam pressure'), it passes a stress flag to the Study module, which reduces the difficulty and duration of the recommended study plan for that day, and to the Fitness module, which substitutes high-intensity workouts with yoga

or walking. Simultaneously, the Finance module checks whether recent overspending in Entertainment is a contributing stressor and, if so, surfaces a financial reflection prompt. This coordinated, multi-domain response to a single emotional signal is beyond the capability of any single-purpose app and represents the core architectural value proposition of YouthMate.

Similarly, the Digital Wellbeing module's detection of Chronic Overuser status triggers a productivity alert in the Study module (recommending the Pomodoro technique and reduced digital study tools), a mental health check-in from the chatbot, and a screen-replacement activity suggestion from the Fitness module. The seamless execution of this cross-module chain within a single user session, with sub-2-second total latency, validates the architectural soundness of the microservice communication design.

C. Performance Bottlenecks and Scaling Considerations

The observed latency degradation at peak load (2.4 seconds with 50+ concurrent users) is the most significant system-level limitation identified in this evaluation. The current deployment relies on a single Docker container instance per microservice; horizontal autoscaling via Kubernetes or AWS Elastic Container Service would address this bottleneck by spawning additional instances under load. Database read performance at scale is also a consideration: MongoDB's aggregation pipeline, used for computing interpretability metrics across multi-session user histories, showed increasing query times beyond 1,000 user records, suggesting that index optimization and data archiving strategies will be required as the user base grows.

D. Limitations

Several limitations of this evaluation should be acknowledged. The performance benchmarks presented are based on simulated or curated test datasets rather than live production user data, which may differ in vocabulary distribution (Mental Health module), spending irregularity (Finance module), and app usage patterns (Digital Wellbeing module). Real-world performance may therefore diverge from reported benchmark figures, particularly for the Mental Health module, where the gap between curated counseling text and the informal, abbreviated language of student chat is substantial.

The SUS evaluation was conducted with a relatively small group of technical evaluators rather than a large, demographically diverse student sample. While the SUS instrument is validated for small-sample administration, a larger-scale usability study would provide more representative and generalizable results. Additionally, the Fitness and Diet module's reliance on user-entered biometric data (rather than wearable sensor integration) introduces a

self-report accuracy concern that limits the module's personalization depth.

E. Future Directions

Three priority enhancement directions emerge from this evaluation. First, real-world data integration through wearable sensors and smartphone APIs would replace simulated and self-entered data with objective physiological and behavioral measurements, substantially improving the accuracy and personalization of the Mental Health, Fitness, and Digital Wellbeing modules. Apple HealthKit, Google Fit, and Android UsageStatsManager APIs are the most viable integration targets.

Second, multilingual NLP support — particularly for Hindi and Marathi in the Indian student context — would dramatically expand YouthMate's accessibility. The current BERT model is fine-tuned only on English text; adapting it to multilingual BERT (mBERT) or a language-specific model such as IndicBERT would address this limitation without requiring full retraining.

Third, federated learning [11] for on-device model training would eliminate the need to transmit sensitive emotional and behavioral data to a central server, addressing the most significant remaining privacy concern. Under federated learning, each user's device trains a local model update on their private data; only the model gradients (not the raw data) are sent to the server for aggregation. This approach would enhance user trust, reduce server-side privacy risk, and comply with emerging data protection regulations including India's Digital Personal Data Protection Act (DPDPA) 2023.

VII. CONCLUSION

This paper presented a comprehensive technical and functional evaluation of YouthMate, a multi-domain AI assistant for youth wellbeing, examining the internal working, algorithmic design, and performance benchmarks of all six AI-powered modules: Mental Health, Study and Career, Finance and Budget, Fitness and Diet, Digital Wellbeing, and Resume and Job Assistance. The evaluation produced four principal findings.

First, all six modules achieved strong individual performance: the BERT-based Mental Health module reached 89.4% emotion classification accuracy; the Study and Career module achieved 86.4% course recommendation accuracy; the Finance regression model demonstrated an MAE of 0.07; and the Resume module achieved 84.9% job-role match accuracy — all consistent with or exceeding comparable single-domain systems in the literature.

Second, system-level performance was validated with average API response latencies of 1.3–1.8 seconds under standard and concurrent load conditions, supported by a 34.6% throughput improvement from the asynchronous

Node.js architecture. Peak load testing identified horizontal autoscaling as the primary infrastructure investment required for production-scale deployment.

Third, the System Usability Scale evaluation yielded an overall mean score of 80.4 (Excellent), confirming that the platform's complexity is effectively abstracted by a well-designed conversational and dashboard interface.

Fourth, and most significantly, the cross-module integration analysis demonstrated that YouthMate's holistic architecture produces emergent intelligence — coordinated multi-domain responses to single user signals — that cannot be replicated by any combination of single-purpose applications. This emergent value is the defining characteristic of YouthMate's contribution to the field of AI-powered youth support.

With planned enhancements including wearable sensor integration, multilingual NLP, federated learning, and production-scale autoscaling, YouthMate is well-positioned to evolve from a validated prototype into a deployable, privacy-preserving, and globally accessible AI ecosystem for holistic student wellbeing.

REFERENCES

- [1] R. Bobade, A. Rajarshi, S. Khot, and P. Sathe, "YouthMate: AI Assistant for Young Minds," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 10, no. 02, Feb. 2026.
- [2] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020.
- [3] E. Kochmar et al., "Automated personalized feedback improves learning gains in an intelligent tutoring system," in *Proc. Int. Conf. on Artificial Intelligence in Education (AIED)*, 2021, pp. 140–146.
- [4] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan et al., Eds. London: Taylor and Francis, 1996, pp. 189–194.
- [5] M. R. Chowdhury, T. Qu, A. S. Aung, S. A. Kaseb, and K. N. Truong, "Predicting and Understanding College Student Mental Health with Interpretable Machine Learning," in *Proc. ACM/IEEE Int. Conf. on Connected Health (CHASE)*, 2025.
- [6] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [7] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study," *Journal of Educational Technology and Society*, vol. 18, no. 3, pp. 75–88, 2015.
- [8] S. K. Jaiswal and P. Tiwari, "AI-Based Emotional Health Monitoring for Students Using Natural Language

Processing," Int. Journal of Artificial Intelligence and Applications, vol. 14, no. 2, pp. 67–78, 2024.

[9] M. A. Bari, S. S. Islam, and M. S. Rahman, "AI-Powered Digital Wellbeing Applications: A Review of Machine Learning Techniques and Challenges," IEEE Access, vol. 11, pp. 11642–11657, 2023.

[10] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL, pp. 4171–4186, 2019.

[11] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.

[12] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in Proc. 31st Int. Conf. on Neural Information Processing Systems (NIPS), pp. 4765–4774, 2017.

[13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.