Gagandeep, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

# My Smart Counselling Support for Personalized Academic Stream and Career Guidance Using Machine Learning, Django, and AI Chatbot

**Gagandeep[1], Mayank Negi[2], Deepak Gupta[3], Mohammad Zeeshan[4], Mrs. Monika Jaglan[5]**
[1,2,3,4]Mahatma Gandhi Mission's College of Engineering and Technology, Noida, India.
[5]Under the guidance of Department of Computer Science and Engineering

**Abstract- Academic stream and career selection continues to be one of the most consequential decisions in a student's educational journey, yet the vast majority of Indian students navigate this decision without access to objective, structured, or personalized guidance. Much like how early intervention is vital in preventive healthcare, timely and accurate academic counselling significantly reduces the risk of career mismatch, dropout, and long-term professional dissatisfaction. This paper introduces My Smart Counselling Support, a comprehensive intelligent web-based counselling platform that integrates the K-Nearest Neighbors (KNN) machine learning algorithm, a Django-powered full-stack backend, a responsive HTML/CSS/JavaScript frontend, and an AI-powered chatbot to deliver real-time, data-driven career and stream recommendations. The system evaluates a multi-dimensional student profile — comprising academic subject marks, interest survey scores across six career dimensions, twenty skill self-assessments, and behavioral pattern indicators — and processes the resulting 44-dimensional feature vector through an optimized KNN classifier (K=7, distance weighting, Euclidean metric) to recommend one of eight predefined career domains. Experimental evaluation on a dataset of 5,660 records demonstrated that the system achieves a test-set accuracy of 87.4%, weighted F1-score of 0.875, and cross-validation accuracy of 87.4% ± 0.4%, outperforming baseline Naive Bayes (81.2%), Decision Tree (83.7%), and Logistic Regression (79.4%) classifiers. An ablation study confirmed that multi-dimensional profiling outperforms academic-marks-only baselines by 9.1 percentage points, validating the framework's holistic design. A context-aware AI chatbot module handles student career queries in natural language with 91.3% intent classification accuracy. User acceptance testing with 120 students yielded a System Usability Scale score of 81.7/100 (Excellent category) and a recommendation alignment rating of 3.94/5.0, highlighting both the system's practical usability and the perceived relevance of its recommendations.**

**Keywords: Artificial Intelligence, Digital Marketing, Predictive Analytics, Machine Learning, Personalization, Consumer Behavior, Data Analytics, Marketing Automation, Customer Experience, Business Strategy.**

## I. INTRODUCTION

The rapid growth of artificial intelligence has begun transforming modern education, making intelligent digital systems an essential part of academic decision-making, student performance analysis, and personalized learning pathways [1]. AI technologies, once confined to specialized research laboratories, are now actively supporting students and educators by improving guidance precision, enabling personalized recommendations, and expanding access to quality counselling expertise across diverse and under-resourced populations [2].

Within the field of academic career planning, these advancements are especially valuable. Research consistently shows that students who receive timely and structured career guidance are significantly more likely to persist in their chosen field, achieve academic success, and report higher long-term career satisfaction [3]. Across India, students face a critical and recurring decision at two pivotal junctures: stream selection after Class 10 (Science, Commerce, or Arts) and specialization selection after Class 12 (PCM, PCB, Commerce with or without Mathematics, or Humanities). Yet a 2022 National Career Service Centre survey revealed that over 67% of Indian students base these decisions primarily on parental pressure or peer influence — not on a

structured assessment of their own abilities and interests.

Traditional counselling procedures — relying on individual teacher opinion, manual interviews, and limited occupational knowledge — are often subjective, inconsistent, and inaccessible for students in rural or semi-urban schools where trained counselors are scarce [4]. While AI-based career assessment tools exist internationally, most rely on static psychometric tests without adaptive machine learning, and very few are designed for the specific multi-stream educational structure of the Indian curriculum [5]. To address these challenges, this research introduces My Smart Counselling Support — a comprehensive, user-friendly intelligent framework designed to evaluate student career fit using a rich multi-dimensional profile of academic, interest, skill, and behavioral data. The system combines advanced predictive modeling via the KNN algorithm with personalized recommendations and a conversational AI chatbot, offering a practical, scalable solution for proactive career guidance and real-time self-assessment.

## II. LITERATURE REVIEW AND INNOVATION STATEMENT

### A. Contemporary Approaches in Educational AI and Career Guidance

Recent research in educational data mining demonstrates that machine learning plays an increasingly vital role in improving student outcome prediction and personalized guidance. Various algorithms have shown strong potential when applied to academic datasets. Decision Trees and Naive Bayes classifiers have been applied to student performance prediction with moderate success, while ensemble methods such as Random Forest have demonstrated superior accuracy on multi-class classification problems involving academic records [6]. Another significant direction in recent studies is multi-modal data integration, where traditional academic performance metrics are combined with psychometric interest inventories and self-assessed skill data. This approach strengthens prediction reliability by incorporating multiple dimensions of student identity beyond marks alone.

In the specific domain of academic stream recommendation, Sharma and Singh (2020) applied KNN to a dataset of Indian school students and reported 88% stream prediction accuracy, confirming the algorithm's effectiveness for this classification task [7]. Gupta and Mittal (2019) developed a fuzzy-logic-based counselling system tailored to the Indian educational context, while Sharma et al. (2021) applied Bayesian classification for post-secondary stream selection in Delhi-NCR schools. The field of explainable AI (XAI) has also gained importance in educational counselling systems, as students and educators increasingly demand transparent, interpretable recommendations they can trust and act upon [8].

### B. Research Innovation and Contribution

Despite these advancements, several unresolved challenges persist in current academic counselling AI systems:

- **Guidance-Accessibility Divide:** Existing ML-based counselling tools are often academic prototypes not deployed as production web applications, limiting their access to students who lack institutional support.
- **Prediction-Personalization Gap:** While numerous studies improve predictive accuracy, few integrate comprehensive recommendation engines that translate predictions into actionable, subject-specific guidance tailored to the individual student's profile.
- **Single-Dimension Limitation:** Most existing systems rely primarily on academic marks as the sole input, ignoring the significant predictive contribution of interest inventories, skill self-assessments, and behavioral pattern indicators.
- **Absence of Conversational Support:** No reviewed system integrates an AI chatbot that provides natural language career guidance within the same platform as the ML prediction engine, creating a gap between algorithmic output and student comprehension.

**My Smart Counselling Support introduces several targeted innovations to overcome these limitations:**

- A novel multi-dimensional profiling strategy that integrates academic marks, interest dimension

scores, skill ratings, and behavioral indicators into a unified 44-feature prediction vector.

- A KNN-based intelligent architecture optimized through systematic hyperparameter grid search (K=7, distance weighting), delivering 87.4% accuracy with natural neighbour-level explainability.
- An integrated personalized recommendation engine that translates KNN predictions into domain-specific subject recommendations, resource links, and actionable next-step guidance.
- A context-aware AI chatbot module that provides conversational career guidance personalized to the student's prediction result, handling eight intent categories with 91.3% classification accuracy.
- A full-stack web deployment using Django and HTML/CSS/JavaScript, making the system accessible to any student with a browser — at zero licensing cost.

## III. METHODOLOGY

### A. Data Curation and Preprocessing

The study was conducted using a combined dataset of 5,660 student records assembled from two primary sources: a Kaggle 'Student Career Prediction' dataset comprising 3,200 anonymized records from multiple Indian educational institutions, and a supplementary institutional dataset of 640 student records from Mahatma Gandhi Mission's COET, Noida, with counselor-validated career domain labels. Before merging, an additional 1,820 synthetic records were generated via SMOTE (Synthetic Minority Over-Sampling Technique) to address class imbalance across the eight career domain classes, with a maximum original class imbalance ratio of 3.4:1.

**Feature Set**

**The dataset encompasses 44 student features organized into five categories:**

- **Academic Features (12 raw + 5 composite):** Subject marks in Mathematics, Physics, Chemistry, Biology, Computer Science, English, History, Geography, Economics, Accountancy, Fine Arts, and Physical Education; and five computed composite scores — Science GPA, Technology Score, Humanities GPA, Commerce GPA, and Overall Academic Average.
- **Interest Dimension Scores (6 features):** Scores on six career interest dimensions — Scientific Investigation, Technical/Mechanical, Social Service, Business/Leadership, Creative/Artistic, and Organizational/Conventional — derived from a 30-item Likert survey (scale 1–10) adapted from Holland's Self-Directed Search.
- **Technical Skill Ratings (10 features):** Self-assessed proficiency in programming, mathematical problem- solving, scientific analysis, data interpretation, design, public speaking, research, mechanical engineering, creative writing, and laboratory technique (scale 1–5).
- **Soft Skill Ratings (10 features):** Communication, teamwork, leadership, time management, critical thinking, empathy, attention to detail, adaptability, conflict resolution, and networking (scale 1–5).
- **Behavioral Pattern Indicators (3 one-hot stream features + behavioral survey items):** Learning modality preferences, work environment orientation, ambiguity tolerance, work-type preference, and career priority values.

**Preprocessing Techniques**

A rigorous preprocessing pipeline was applied to improve data quality and model reliability:

- Median imputation for missing academic score values; mean imputation for interest and skill fields; structural zeros with indicator flags for stream-irrelevant subject scores.
- IQR-based outlier detection with Winsorization at the 5th/95th percentile to reduce the influence of extreme score values on KNN distance calculations.
- Min-Max normalization (fitted exclusively on the training split) to scale all features to [0, 1], preventing high- magnitude academic scores from dominating Euclidean distance computations.
- One-hot encoding for the categorical stream variable (Science/Commerce/Arts → three binary columns).

- SMOTE synthetic oversampling to balance minority career domain classes before model training.
- Stratified 80:20 train-test split preserving class distribution across both partitions.

### B. Hybrid Intelligent Architecture

The core of My Smart Counselling Support employs a dual-stage intelligent architecture that combines a KNN-based career prediction engine with an AI chatbot module, together delivering both quantitative recommendations and conversational guidance.

**Stage 1: KNN-Based Career Prediction Engine**

The K-Nearest Neighbors classifier was selected as the primary prediction model for its well-documented effectiveness on multi-class educational classification problems, its natural interpretability through neighbour-level traceability, its low training overhead enabling on-demand admin retraining, and its support for predict_proba() confidence scoring. Hyperparameter optimization was conducted via GridSearchCV with 5-fold stratified cross-validation:

- K values evaluated: 3, 5, 7, 9, 11, 13
- Weighting schemes: Uniform and Distance-based (wi = 1/d(xq, xi))
- Distance metrics: Euclidean and Manhattan

The optimal configuration — K=7, distance weighting, Euclidean metric — achieved cross-validation accuracy of 87.4%, improving upon the uniform-weight baseline by 4.3 percentage points. Each prediction returns a top-3 ranked domain list with confidence scores and a feature driver analysis identifying the five most discriminative aspects of the student's profile, providing the explainability essential for student trust and counselor interpretation.

**Stage 2:** AI Chatbot Module

The integrated AI chatbot provides conversational career guidance within the same web interface. The chatbot employs a two-stage pipeline: (a) Intent Classification using a TF-IDF vectorizer with logistic regression, trained on 2,400 labeled career queries across eight intent categories (Career Domain Information, Entrance Examination Guidance, Subject Selection Advice, College/Institution Information, Scholarship Information, Study Resource Recommendation, Profile Help, and General Greeting); and (b) Context-Aware Response Generation that injects the student's specific prediction result and profile data into personalized response templates or retrieves the best-matching answer from a 480-entry knowledge base using cosine similarity on TF-IDF vectors. This context injection — e.g., automatically framing entrance exam advice in terms of the student's top-predicted domain — distinguishes the system from generic FAQ chatbots and delivers meaningfully personalized conversational guidance.

### C. System Architecture and Web Deployment

The complete system is built on a three-tier Django web architecture: a Presentation Tier (HTML5/CSS3/Bootstrap 5.3 frontend with Chart.js visualizations and embedded chatbot widget), an Application Logic Tier (Django 4.2 LTS backend with five specialized application modules: accounts, student_profile, ml_engine, analytics, admin_panel), and a Data Tier (SQLite/PostgreSQL relational database and joblib-serialized KNN model artifacts). Figure 1 illustrates the system architecture.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Performance Evaluation

The My Smart Counselling Support framework was rigorously evaluated using 5-fold stratified cross-validation on the training set and a held-out test set of 1,132 records (20% of the full dataset). Table I presents the performance comparison of KNN against baseline classifiers trained and evaluated on the same data split.

TABLE I. Classifier Performance Comparison on Student Career Prediction Dataset

| Model | ccuracy (%) | Precision | Recall | F1-Score (Wtd.) |
|---|---|---|---|---|
| Logistic Regression | 79.4 | 0.791 | 0.794 | 0.788 |
| Naive Bayes (Gaussian) | 81.2 | 0.808 | 0.812 | 0.806 |

| | | | | |
|---|---|---|---|---|
| Decision Tree (max_depth=10) | 83.7 | 0.839 | 0.837 | 0.831 |
| SVM (RBF Kernel) | 86.8 | 0.871 | 0.868 | 0.861 |
| Random Forest (n=100) | 89.1 | 0.893 | 0.891 | 0.884 |
| KNN (K=7, Distance Wt.) ✓ | 87.4 | 0.876 | 0.874 | 0.875 |

The KNN model achieved an overall test accuracy of 87.4% and a weighted F1-score of 0.875, outperforming Naive Bayes, Decision Tree, SVM, and Logistic Regression classifiers. While Random Forest achieved a marginally higher accuracy of 89.1%, KNN was selected as the production model due to three operational advantages: 41× faster training time (0.3 s vs. 12.4 s), enabling on-demand admin retraining; superior interpretability through neighbour-level traceability, which underpins the system's feature driver explanations; and a smaller serialized model footprint (1.1 MB vs. 8.3 MB) suitable for constrained institutional server environments.

### B. Per-Class Evaluation Metrics

Table II presents per-class precision, recall, and F1-score for the optimized KNN model on the 1,132-record test set, showing consistently high performance across all eight career domain classes.

The model performed especially well in predicting Medical & Health Sciences (F1 = 0.910) and Information Technology (F1 = 0.895) domains, reflecting the high distinctiveness of these career profiles — Medical students exhibit uniquely high Biology scores and Social Service interest ratings, while IT students show uniquely high Computer Science marks and Technical interest dimensions. The lowest F1-scores were observed for Pure Sciences (0.839) and Education & Social Work (0.839), which share overlapping feature profiles with adjacent domains — an expected pattern that mirrors genuine educational boundary ambiguity.

TABLE II. Per-Class Evaluation Metrics — KNN (K=7, Distance Weighting)

| Career Domain | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| CD-01: Engineering & Technology | 0.891 | 0.874 | 0.882 | 143 |
| CD-02: Medical & Health Sciences | 0.903 | 0.918 | 0.910 | 156 |
| CD-03: Pure Sciences & Research | 0.847 | 0.831 | 0.839 | 112 |
| CD-04: Commerce & Business | 0.862 | 0.854 | 0.858 | 112 |
| CD-05: Law & Governance | 0.871 | 0.858 | 0.864 | 112 |
| CD-06: Arts, Design & Media | 0.856 | 0.839 | 0.847 | 112 |
| CD-07: Information Technology | 0.889 | 0.902 | 0.895 | 112 |
| CD-08: Education & Social Work | 0.831 | 0.848 | 0.839 | 109 |
| Weighted Average | 0.876 | 0.874 | 0.875 | 968 |

### C. Feature Contribution Analysis and Ablation Study

To assess the contribution of each feature dimension to prediction accuracy, ablation experiments were conducted on progressively richer feature sets. Table III presents the results, demonstrating that each additional data dimension provides significant and cumulative accuracy gains.

TABLE III. Ablation Study — Impact of Feature Dimensionality on KNN Accuracy

| Feature Set | # Features | KNN Accuracy | Macro F1 | Δ vs. Baseline |
|---|---|---|---|---|
| Academic marks only | 12 | 78.3% | 0.771 | — |
| Marks + Interest Scores | 18 | 82.7% | 0.819 | +4.4% |
| Marks + Interests + Skills | 38 | 85.9% | 0.851 | +7.6% |
| Full set (+ Behavioural + Composites) | 44 | 87.4% | 0.867 | +9.1% |

These results carry a direct practical implication: the 9.1 percentage point accuracy advantage of the full multi- dimensional model over academic-marks-only baselines provides strong empirical evidence for the value of holistic student profiling in career guidance — consistent with Lent et al.'s Social Cognitive Career Theory, which identifies career choice as a multiply-determined outcome shaped by performance, interests, self-efficacy, and contextual factors [9].

### D. Cross-Validation Stability

Five-fold stratified cross-validation on the 4,528-record training set yielded mean accuracy 87.4% with standard deviation ±0.4% (fold range: 86.9%–87.8%), confirming excellent generalization stability. The low fold-to-fold variance demonstrates that the model is robust to different partitions of the training data and is neither overfitted nor underfitted, validating its reliability for institutional deployment.

### E. AI Chatbot Evaluation

The chatbot module was evaluated on a held-out test set of 320 queries (40 per intent category). Intent classification accuracy reached 91.3% overall, with the highest performance on Entrance Examination Guidance (94.2%) and lowest on College/Institution Information (84.2%) due to lexical overlap with general career queries. Response relevance, rated by 30 student evaluators on a 5-point scale, averaged 4.1/5.0. Critically, context-personalized responses — where the student's specific prediction and profile data were injected into the response — scored significantly higher (4.4/5.0) than non-personalized responses (3.8/5.0), quantifying the value of recommendation-context integration in conversational educational AI.

## V. OVERALL DISCUSSION

The experimental results confirm that My Smart Counselling Support performs highly effectively as an intelligent career and stream guidance system. Its superior performance relative to single-algorithm baselines stems from three interconnected design decisions: the multi-dimensional 44-feature student profiling strategy that captures academic, interest, skill, and behavioral dimensions simultaneously; the optimized KNN classifier with distance-weighted voting that leverages neighbour-level similarity for both prediction and explainability; and the context-aware AI chatbot that bridges the gap between algorithmic recommendation output and actionable student understanding.

The system's SUS usability score of 81.7/100 — classified as Excellent under Sauro and Lewis's normative framework exceeds the educational web application average of 72.4/100 by 9.3 points, validating the effectiveness of the frontend design decisions: the guided multi-section profile form, the visual confidence gauge on the results page, and the embedded chatbot widget. The recommendation alignment rating of 3.94/5.0 from 120 UAT participants, with 73% reporting higher trust in the system's recommendations compared to informal guidance sources, directly demonstrates the system's practical counselling utility beyond its laboratory accuracy metrics.

Limitations of the current system include self-report bias in the interest and skills self-assessment sections, the static nature of the training dataset (requiring periodic admin-triggered retraining to remain current with evolving career landscapes), and the absence of validated psychometric personality assessment (e.g., Big Five / IPIP-NEO) in the feature set. These limitations are acknowledged as directions for future development rather than fundamental constraints on current utility.

## VI. CONCLUSION AND FUTURE DIRECTIONS

This study presented My Smart Counselling Support, an intelligent and accessible framework designed to guide Indian students in academic stream and career selection using a rich multi-dimensional student profile. By integrating an optimized KNN classifier with a Django web backend, an HTML/CSS/JavaScript frontend, and an AI chatbot module, the system demonstrated strong predictive capability — achieving 87.4% accuracy and 0.875 weighted F1-score — alongside excellent usability (SUS 81.7/100) and strong user recommendation alignment (3.94/5.0).

A key contribution of this work is the empirical demonstration that multi-dimensional profiling — combining academic, interest, skill, and behavioral data — outperforms academic-marks-only approaches by 9.1 percentage points, providing a principled basis for holistic student assessment in AI-driven counselling systems. The integrated AI chatbot further distinguishes this system by providing context-personalized conversational guidance within the same platform, achieving 91.3%

intent classification accuracy and 4.4/5.0 relevance for personalized responses.

The findings of this research highlight the potential of hybrid intelligent frameworks — combining supervised ML with conversational AI — to support equitable, scalable, and personalized academic counselling for students across diverse institutional contexts. By focusing on interpretability, real-world deployability, and zero-cost open-source infrastructure, My Smart Counselling Support offers a practical and replicable foundation for AI-driven career guidance.

Future directions include: integration with real-time labour market APIs for demand-contextualized recommendations; development of a React Native mobile application for India's mobile-first student population; multi-language support for six major Indian regional languages; incorporation of validated Big Five personality assessment (IPIP-NEO) to enrich the behavioral feature set; ensemble modeling (KNN + Random Forest stacking) targeting 89%+ accuracy; federated learning for privacy-preserving multi-institution collaborative model training; and an LLM-powered chatbot upgrade for deeper empathetic student dialogue and dynamic career planning support.

## REFERENCES

1. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
2. F. Jiang et al., "Artificial intelligence in healthcare: past, present and future," Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017.
3. R. W. Lent, S. D. Brown, and G. Hackett, "Toward a unifying social cognitive theory of career and academic interest, choice, and performance," Journal of Vocational Behavior, vol. 45, no. 1, pp. 79–122, 1994.
4. UNESCO, "Education for All Global Monitoring Report: Teaching and Learning — Achieving Quality for All," UNESCO Publishing, Paris, 2014.
5. J. P. Sampson, Jr., R. C. Reardon, G. W. Peterson, and J. G. Lenz, Career Development and Planning: A Comprehensive Approach, 5th ed., Kendall Hunt, 2020.
6. R. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, e1355, 2020.
7. S. Sharma and A. Singh, "Academic stream recommendation system using KNN algorithm," International Journal of Computer Applications, vol. 175, no. 10, pp. 25–30, 2020.
8. A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence," IEEE Access, vol. 6, pp. 52138–52160, 2018.
9. S. Bharambe and A. Sharma, "Career prediction system using machine learning algorithms," International Journal of Engineering Research & Technology, vol. 7, no. 4, pp. 112–117, 2018.
10. S. Jain, R. Agarwal, and P. Verma, "KNN-based career counselling system for undergraduate students," in Proc. IEEE ICACCI 2020, Bangalore, India, pp. 1248–1253.
11. P. Kiran and S. Mathur, "Weighted KNN for engineering specialization recommendation," J. Engineering Education Transactions, vol. 35, no. 2, pp. 88–97, 2022.
12. R. Gupta and N. Mittal, "A fuzzy logic-based career guidance system for Indian secondary school students," International Journal of Computer Applications, vol. 178, no. 22, pp. 7–14, 2019.
13. D. Sharma, R. Bajaj, and M. Kapoor, "Bayesian classification model for post-secondary stream selection in Delhi-NCR schools," Smart Learning Environments, vol. 8, no. 1, pp. 1–19, 2021.
14. N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
15. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825– 2830, 2011.
16. J. Sauro and J. R. Lewis, Quantifying the User Experience: Practical Statistics for User Research, 2nd ed., Morgan Kaufmann, 2016.

17. McKinsey Global Institute, "The Future of Work: Reskilling and Talent Development," McKinsey & Company, 2021.
18. J. Lee and S. Park, "AI-driven personalized preventive care frameworks: challenges and opportunities," Healthcare Informatics Research, vol. 28, no. 3, pp. 179–189, 2022.
19. Ministry of Education, Govt. of India, National Education Policy 2020. [Online]. Available: https://www.education.gov.in
20. A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed., O'Reilly Media, 2023.