

PAPER

Influencing Factors of Vocational College Students' Career Development: A GA-Optimized KNN-Random Forest Model

Fei Gao¹, Jing Yang²(✉)¹Zhejiang Shuren University,
Hangzhou, China²Hebei Women and
Children Activity Center,
Shijiazhuang, Chinapljinger@126.com**ABSTRACT**

College students can effectively realize personal value and promote social progress by attaching great importance to career planning and development. This study collected 404 valid data points through the career aspiration questionnaire, career exploration questionnaire, and career adaptability questionnaire and used the K-nearest neighbor (KNN) random forest algorithm with hyperparameter optimization based on genetic algorithms (GA) for influencing factor analysis. First, data cleaning was performed, followed by basic frequency analysis and difference analysis. Next, the KNN algorithm was introduced to establish a random forest model, with parameters for the random forest determined through hyperparameter analysis using a genetic algorithm (GA). Subsequently, data classification and value assignment were conducted to predict the importance ranking of influencing factors. Finally, the model's accuracy (Acc), precision (P), recall (R), and F1 score were evaluated and compared with the unimproved Random Forest algorithm. Results demonstrated that the KNN Random Forest algorithm significantly improved Acc compared to traditional methods. The study found that in terms of career aspiration and adaptability, the top three factors were participation in part-time jobs, internships, or social practices; grade; and career counseling experience. For career exploration, the top three factors were grade, participation in part-time jobs, internships, or social practices, and first-generation college student. Additionally, there were variations in the importance rankings of each dimension in career aspiration, career exploration, and career adaptability. Higher vocational colleges need to further strengthen guidance on internships and practical training, as well as management of career counseling, to more effectively promote students' career development.

KEYWORDS

career development, K-nearest neighbor (KNN), random forest algorithm, genetic algorithm (GA), vocational college students

1 INTRODUCTION

College years represent an important stage in career development, crucial for forming career identity, clarifying future goals, and accumulating knowledge

Gao, F., Yang, J. (2026). Influencing Factors of Vocational College Students' Career Development: A GA-Optimized KNN-Random Forest Model. *International Journal of Interactive Mobile Technologies (IJIM)*, 20(13), pp. 137–153. <https://doi.org/10.3991/ijim.v20i13.62394>

Article submitted 2026-03-20. Revision uploaded 2026-05-17. Final acceptance 2026-05-20.

© 2026 by the authors of this article. Published under CC-BY.

and expertise while enhancing comprehensive competencies. Career aspiration, career exploration, and career adaptability can measure students' readiness and progress. Career aspiration reflects personal ambitions and expectations, serving as self-awareness of professional perceptions and forming the foundation for career development. Career exploration, involving understanding, experimentation, and evaluation, constitutes the critical action in career development. Career adaptability, as a self-regulation ability in human-environment interactions, demonstrates overall competence. Analyzing the influencing factors of these three aspects holds significant importance for better promoting college students' career development [1, 2].

Career aspirations are typically conceptualized as an individual's goals for their future profession. Regarding the impact of gender, York [3] pointed out that some female valedictorians did intend to have similar college and career outcomes to the males, but other females tended to lower educational aspirations and lower-paying careers. Nadeem and Khalid [4] found that males tend to be more conservative in career aspiration and exhibit lower levels compared to females. Research by Chang et al. [5] revealed varying levels of career aspirations among students across different grades. Al-Bahrani et al. [6] demonstrated that parental education levels significantly influence adolescents' career aspirations.

Zikic and Klehe [7] defined career exploration as collecting information related to personal career development. Chen et al. [8] believed career exploration was conducive to self-reflection and self-regulation. Regarding influencing factors, there is still considerable debate among scholars regarding the role of gender differences. Stumpf and Lockhart [9] noted that gender was also related to beliefs regarding the instrumentality of exploration. Zhang and Ye [10] observed that though male and female adolescents scored differently across various dimensions of career exploration—such as males scoring higher than females in environment exploration and females scoring higher than males in self-exploration—these differences did not reach statistical significance. Wang et al. [11] identified significant gender disparities in the career exploration dimension of career construction behavior. Regarding family factors, Kracke [12] investigated how parental behaviors influence adolescents' career exploration. The findings indicated that child-centered, supportive, and reciprocal parent-adolescent relationships correlate with more active exploration. In terms of age, Taveira et al. [13] noted that grade-level was one of the significant and prevailing individual characteristics related to adolescents' career exploration.

In terms of career adaptability, Savickas and Porfeli [14] pointed out that the ever-changing professional environment requires individuals to promptly adjust their state and strive for harmony between their personal condition and external circumstances. Regarding influencing factors of career adaptability, Zhao and Guo [15] noted that differences in traditional social expectations for gender roles lead to varying levels of career adaptability between males and females. However, some studies suggested that gender didn't affect career adaptability. Hirschi's research on middle school students found gender didn't affect career adaptation [16]. Olugbade [17] discovered that age differences led to variations in career adaptability, with higher age corresponding to greater adaptability. Xiao [18] indicated that there were significant grade-level differences in college students' career adaptability. Zulfiani and Khaerani [19] identified family support as one of the influencing factors in students' career adaptability. Liu and Zhang [20] revealed that social support was associated with higher career adaptability indirectly.

However, existing research on factors influencing career aspiration, career exploration, and career adaptability requires further enrichment. With the advancement of computer technology, machine learning has enabled more in-depth analysis of various issues. For instance, methods such as Random Forest have been widely

adopted in the field of education. For example, Kiu [21] and Pellagatti et al. [22] employed this method to analyze educational issues.

Although ensemble learning algorithms such as Random Forest have demonstrated good performance on most datasets, there is still room for further improvement in their prediction accuracy (Acc). The limitation of the Random Forest Algorithm lies in its reliance solely on majority voting to determine results, while the shortcoming of the K-Nearest Neighbor (KNN) algorithm is that it does not account for the differences in feature importance. This paper chooses the Random Forest algorithm because it is one of the best-performing algorithms in the field of machine learning; the KNN algorithm is selected due to its simple and intuitive logical framework. By optimizing the KNN algorithm and embedding it into the Random Forest algorithm for improvement, this study adopts a KNN-improved Random Forest algorithm to predict the importance of influencing factors of career aspiration, career exploration, and career adaptability and conducts a comparative analysis before and after algorithmic improvements. Among these, career aspiration is measured using the Chinese version of CAS-R revised by Shi Jian. CAS-R, developed by Gregor and O'Brien [23], includes 24 items across three dimensions: achievement aspiration, leadership aspiration, and educational aspiration. Career exploration is assessed using the Stumpf et al. questionnaire revised by Xu [24], consisting of 18 items across four dimensions: environment exploration, self-exploration, purpose-systematic exploration, and amount of information. Career adaptability is measured using the CAAS China version revised by Hou et al. [25], which includes 24 items across four dimensions: concern, control, curiosity, and confidence.

2 DATA PROCESSING

2.1 Data cleaning

The original sample dataset contains 425 basic data entries, each comprising 75 field variables. Prior to model construction, the data undergoes cleaning to correct incomplete, inconsistent, and noisy data.

1. Processing of missing data. No missing data was identified upon inspection.
2. Checking for outliers. Exclude data entries with all answers identical (20 entries). Exclude entries with students listed as majors (1 entry).
3. Removing unconsidered fields. This includes "Serial Number," "College Name," and "Major Name."
4. Implementing standardized naming.

After processing, 404 pieces of data were obtained.

2.2 Frequency analysis

Descriptive statistics are used to analyze the demographic variables of 404 cases. Table 1 presents the basic characteristics of the data, while Table 2 shows the demographic scores for career aspirations, career exploration, and career adaptability. Specifically, males score higher than females on career aspirations, while females score higher than males on career exploration and career adaptability. Students who held student leadership positions score higher than those who did not on all three dimensions. Only children in the family score higher on career exploration than non-only children but score lower on both career aspirations and career adaptability. The first-generation college students in their families score higher on career exploration

and career adaptability than non-first-generation college students but score lower on career aspirations. Students residing in urban areas score higher on all three dimensions than those residing in rural areas. Students who receive career counseling score higher on all three dimensions than those who did not. Career aspirations and career exploration increase with higher grade levels, while career adaptability initially decreases and then increases. In all three dimensions, students who participate in part-time jobs, internships, and social practice related to their major score higher than those who participate in activities unrelated to their major or do not participate in any activities.

Table 1. Frequency analysis of demographic variables

Variable	Category	Frequency	Percentage	Variable	Category	Frequency	Percentage
Gender	Male	186	46.0	First-generation college student	Yes	218	54.0
	Female	218	54.0		No	186	46.0
Grade	Freshman	199	49.3	Part-time jobs, internships or social practices	Yes, professionally related	157	38.9
	Sophomore	83	20.5		Yes, professionally unrelated	165	40.8
	Junior	122	30.2		No	82	20.3
Student leader	Yes	116	28.7	Career counseling	Yes	212	52.5
	No	288	71.3		No	192	47.5
Only child	Yes	115	28.5	Family residence	Urban area	161	39.9
	No	289	71.5		Rural area	243	60.1

Table 2. Demographic differences in career aspiration, career exploration, and career adaptability

Variable	Category	Career Aspirations	Career Exploration	Career Adaptability
Gender	Male	2.36 ± 0.77	3.22 ± 0.77	3.61 ± 0.67
	Female	2.22 ± 0.74	3.23 ± 0.72	3.66 ± 0.6
Student leaders	Yes	2.49 ± 0.65	3.45 ± 0.7	3.84 ± 0.64
	No	2.2 ± 0.78	3.14 ± 0.75	3.55 ± 0.62
Only child	Yes	2.26 ± 0.76	3.26 ± 0.8	3.53 ± 0.65
	No	2.29 ± 0.75	3.21 ± 0.72	3.68 ± 0.63
First-generation college student	Yes	2.27 ± 0.79	3.25 ± 0.76	3.64 ± 0.66
	No	2.29 ± 0.71	3.2 ± 0.73	3.63 ± 0.6
Family residence	Urban area	2.35 ± 0.78	3.36 ± 0.74	3.69 ± 0.62
	Rural area	2.24 ± 0.74	3.14 ± 0.74	3.6 ± 0.64
Career counseling	Yes	2.32 ± 0.69	3.35 ± 0.73	3.7 ± 0.63
	No	2.24 ± 0.82	3.1 ± 0.74	3.56 ± 0.63
Grade	Freshman	2.24 ± 0.81	3.15 ± 0.74	3.63 ± 0.62
	Sophomore	2.29 ± 0.77	3.24 ± 0.78	3.57 ± 0.65
	Junior	2.35 ± 0.64	3.34 ± 0.71	3.69 ± 0.64
Part-time jobs, internships, or social practices	Yes, professionally related	2.35 ± 0.79	3.36 ± 0.81	3.74 ± 0.67
	Yes, professionally unrelated	2.27 ± 0.69	3.22 ± 0.64	3.62 ± 0.59
	No	2.16 ± 0.81	2.99 ± 0.77	3.48 ± 0.63

3 MODEL ESTABLISHMENT

3.1 KNN random forest algorithm

The Random Forest algorithm demonstrates high Acc, interpretability, and adaptability, excelling in solving multi-feature problems by constructing multiple weak decision trees and synthesizing their prediction results. The process includes the following steps: ① Sampling: Use the bootstrap method to sample the original dataset, generating multiple subsets. ② Feature selection: For each subset, randomly select a subset of features as candidate features. ③ Building decision trees: For each subset, use the selected features to build a decision tree. At each node, select the optimal splitting feature to maximize classification Acc. ④ Combination: Combine all the constructed decision trees into a forest to form a powerful classifier. The KNN algorithm, as a supervised learning method, can be applied to both classification and regression tasks. Its core strengths include logical simplicity, robustness, and strong adaptability to noisy data. However, KNN also has limitations, such as high computational cost, memory-intensive operation, and no clear standard for K value selection, leading to the widespread adoption of improved KNN variants in practice. The core of the KNN algorithm lies in the measurement of distances between samples. This study uses the Manhattan distance as a basis and introduces the range of feature values to perform normalization, thereby eliminating differences in units among different features. The basic distance formula is shown in Equation 1:

$$D_m = \sum_{i=1}^n \text{abs} \frac{(x_i - y_i)}{r_i} \quad (1)$$

In Equation 1, n is the total number of features used for prediction in this study, comprising 8 demographic features and item features from various dimensions of career development; x_i is the value of the i -th demographic feature for the neighbor sample; y_i is the value of the i -th demographic feature for the student sample to be predicted; r_i is the range of values for the i -th demographic feature across the 404 valid samples. However, Equation 1 does not account for differences in feature importance; that is, when calculating the distance between neighboring nodes, it does not distinguish the weights of different features. To address this, we introduce a feature importance factor in Equation 2, resulting in the following improved Manhattan distance formula:

$$D_m = \sum_{i=1}^n \text{abs} \frac{f_i * (x_i - y_i)}{r_i} \quad (2)$$

In Equation 2, f_i is the importance weight of the i -th feature. The K-nearest neighbors algorithm proceeds as follows. First, the range of each feature is calculated, and the i -th demographic feature is normalized. Next, the improved Manhattan distance Equation 2 is used to calculate the distance between the test student sample and its nearest neighbor samples. Finally, after calculating the distances between the test sample and all nearest neighbor samples within the decision tree leaf node, the distance results are stored and sorted in ascending order (from smallest to largest) in a dictionary.

Since all columns in the dataset are categorical variables and the predictor variables are also categorical variables, we use the chi-square test to calculate the chi-square value of the independent variable column relative to the dependent variable column. The feature with the smallest chi-square value is considered to have the greatest impact on the dependent variable. The f_i feature importance weighting score is calculated as follows:

$$f_i = \frac{\text{chiscore}(i)}{\sum_{x=1}^n \text{chiscore}(x)}$$

In the formula above, the importance weight $\text{chiscore}(i)$ of the i -th feature is the chi-square value of the i -th feature. After calculating and storing the distances from the test node to its neighboring nodes, instead of selecting a fixed number of nearest neighbors from the sorted distance results, a 50% threshold rule is applied to select the top 50% of nearest neighbors for majority voting. In this way, the bottom 50% of distant neighbors—which are unsuitable for decision-making—are directly excluded; when executing the final step of the algorithm (i.e., performing majority voting among the qualified nearest neighbors), the model's prediction Acc is improved.

3.2 Genetic algorithm

Improving the predictive performance of the KNN random forest model is influenced by multiple key hyperparameters. As external parameters set prior to model training, hyperparameters cannot be learned autonomously through data training; instead, their optimal combination must be determined through global optimization. Traditional grid search and random search methods suffer from high computational costs and a tendency to get stuck in local optima when dealing with scenarios involving multiple parameters and a wide search range. To address this, this study introduces a GA to achieve efficient global optimization of hyperparameters. This approach can efficiently handle high-dimensional, multi-constrained hyperparameter spaces without the need to exhaustively search all parameter combinations, thereby significantly improving the efficiency of hyperparameter analysis and meeting the model tuning requirements for the small-sample survey data in this study. GA is a stochastic global search and optimization method developed by mimicking the mechanisms of biological evolution in nature. They treat all individuals within a population as objects and utilize randomization techniques to guide an efficient search of an encoded parameter space. Genetic algorithms (GA) primarily simulate the three core evolutionary operations of “selection, crossover, and mutation,” iterating progressively through complex solution spaces to ultimately converge on a global optimal solution.

- 1. Core hyperparameters and optimization range settings:** Based on the characteristics of the classification prediction task for student career development, five core hyperparameters to be optimized are determined, and their meanings, functions, and optimization ranges are shown in Table 3.

Table 3. Meanings, functions, and optimization ranges of five core hyperparameters

Parameter Name	Type	Meaning and Function	Optimization Range
n_trees	integer	The number of decision trees in a random forest. The more trees there are, the better the ensemble performance and the lower the variance; however, computational costs increase accordingly, so a balance must be struck between Acc and efficiency.	50–500
max_depth	integer	The maximum depth of a single decision tree, which controls model complexity to prevent overfitting. Excessive depth tends to fit noise, while insufficient depth easily causes underfitting.	5–80
min_samples_split	integer	The minimum number of samples required for a decision tree node split. The larger the value, the more conservative the model, which helps prevent overfitting caused by excessive splitting.	2–10
min_samples_leaf	integer	The minimum number of samples a leaf node must contain. Increasing this parameter helps prevent the generation of overly fine-grained leaf nodes and accommodates the characteristics of small-sample datasets.	1–5
n_neighbors	integer	Improving the value of K in the KNN algorithm directly affects the effectiveness of local decision-making within leaf nodes.	3–15

The primary objective of hyperparameter optimization is to identify, within the aforementioned parameter space, the combination of parameters that yields the best overall performance for the model across all dimensions of career development prediction. This study adopts the model's average classification Acc as the core optimization objective, while also taking into account the model's generalization ability and computational efficiency, thereby providing a solid foundation for constructing the fitness function of the genetic algorithm.

2. Adaptive design of GA hyperparameter optimization: This study employs a decimal encoding scheme to sequentially combine the five hyperparameters to be optimized in Table 3, encoding them as a 5-digit decimal chromosome. Each gene position corresponds to the value of a hyperparameter. This approach eliminates the need for conversion between binary and decimal formats, resulting in higher computational efficiency and clearer physical interpretation. The values of the fitness function directly quantify the relative quality of each hyperparameter combination. This study constructs a fitness function with the objective of maximizing the model's average classification Acc.

$$\max_{\theta \in \Theta} f(\theta) = \frac{\sum_{i=1}^n Acc_i}{n}$$

In the formula above, θ represents a set of hyperparameter combinations, i.e., a chromosome consisting of five hyperparameters; n is the total number of career development dimensions to be predicted in this study. $Acc(\theta)$ denotes the model's prediction Acc for the i -th career development dimension under the hyperparameter combination θ . $f\theta$ represents the fitness value corresponding to this hyperparameter combination, with a range of [0, 1], where a higher value indicates superior

overall predictive performance of the hyperparameter combination. Considering the scale of the hyperparameter space and computational efficiency requirements in this study, the initial population size of the GA is set to 10, with a maximum of 10 generations. The specific genetic operations are designed as follows:

Encoding and initial population generation: each hyperparameter combination is encoded as a “chromosome.” The initial population is generated randomly as a set of randomly generated hyperparameter combinations. The population size is set to 10 to cover the initial solution space of different hyperparameter combinations.

Fitness calculation: For each chromosome (hyperparameter combination) in the population, a modified KNN random forest model is trained to calculate its fitness score. A higher fitness score indicates that the hyperparameter combination has better predictive performance, and thus a higher probability of being retained and passed on.

Selection process: Using a roulette wheel selection method, the probability of an individual being selected is calculated based on its fitness value. This approach prioritizes retaining high-fitness individuals while eliminating low-fitness ones, thereby simulating the principle of “survival of the fittest” in biological evolution.

Crossover operation: For the selected high-quality individuals, the algorithm simulates genetic recombination on these “chromosomes,” exchanging gene segments with a crossover probability of 0.8 to generate new hyperparameter combinations (offspring individuals), thereby enriching the diversity of the solution space. For the selected high-quality individuals, the algorithm simulates genetic recombination on these “chromosomes,” exchanging gene segments with a crossover probability of 0.8 to generate new hyperparameter combinations (offspring individuals), thereby enriching the diversity of the solution space.

Mutation operation: In order to avoid the algorithm getting stuck in local optima and prevent the found hyperparameter combinations from not being globally optimal, random gene mutations are performed on offspring individuals with a low probability of 0.05, and hyperparameter values are randomly adjusted.

Population update: An elite retention strategy is employed, whereby the individual with the highest fitness in the current generation is directly retained for the next generation, while individuals with lower fitness in the current generation are replaced by their offspring, thereby completing the update.

Iteration termination: Through multiple iterations, hyperparameter optimization is completed. Set the maximum number of iterations to 10 generations and set a termination threshold: if the optimal fitness value of the population for five consecutive generations improves by less than 0.001, terminate the iteration early. After the iteration is terminated, output the chromosome with the highest fitness in the population, which is the optimal hyperparameter combination of the model.

4 RESULT ANALYSIS

4.1 Analysis of influencing factors

Table 4. Key parameters in the KNN Random Forest algorithm

Parameter Name	Type	Value
n_trees	integer	70
max_depth	integer	50
min_samples_split	integer	5
min_samples_leaf	integer	2

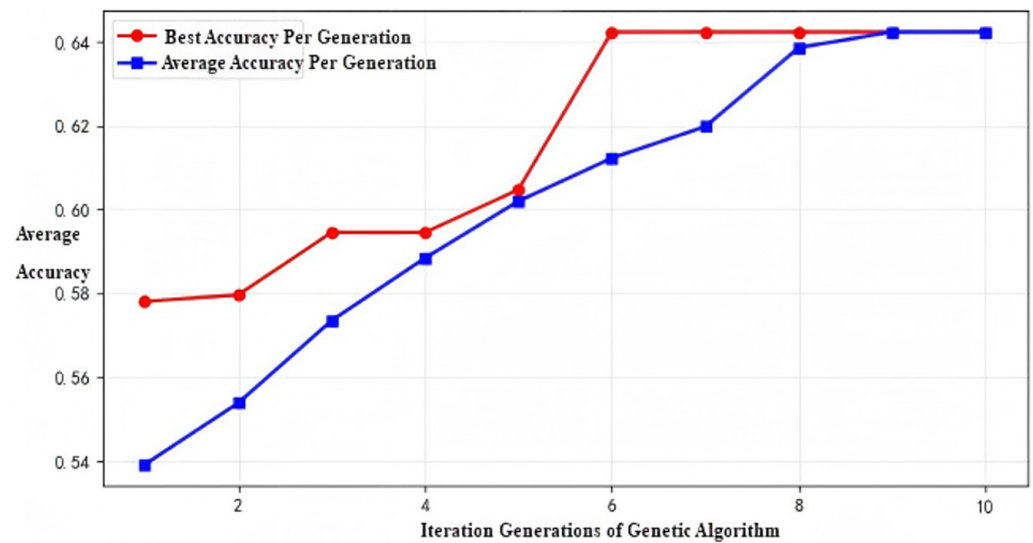


Fig. 1. Acc curve of iteration process in genetic algorithms

Based on the 404 valid questionnaire data collected in this study, the GA converged within 10 generations, as shown in Figure 1, with the population's optimal fitness value (average prediction Acc) gradually increasing and eventually stabilizing. This validates the effectiveness of the GA for hyperparameter optimization of this model, demonstrating its ability to rapidly identify a globally optimal solution. The optimal combination of hyperparameters for the model is shown in Table 4.

Classify the data into four categories based on the quartiles of the numerical distribution, assigning a value of 4 to the bottom 25% and decreasing values sequentially for the subsequent intervals, thereby converting continuous values into categorical values. The Random Forest algorithm is then used for prediction, with variable importance scores to evaluate the impact of each demographic variable on the dependent variable. The results calculated the importance of variables in career aspiration, career exploration, and career adaptability, as shown in Table 5.

Table 5. Ranking of the importance of demographic factors influencing career aspiration, career exploration, and career adaptability

	Gender	Grade	Student Leader	Part-Time Jobs, Internships, or Social Practices
Career aspiration	0.1016	0.1763	0.0574	0.2055
Career exploration	0.1215	0.1663	0.0679	0.1598
Career adaptability	0.0967	0.1856	0.0849	0.1679
Achievement aspiration	0.1132	0.1766	0.0622	0.1804
Leadership aspiration	0.0906	0.1688	0.0802	0.1913
Educational aspiration	0.0961	0.1677	0.0587	0.187
Environment exploration	0.1244	0.1718	0.1057	0.146
Self-exploration	0.1124	0.1848	0.0703	0.1589
Purpose-systematic exploration	0.1234	0.172	0.0649	0.174

(Continued)

Table 5. Ranking of the importance of demographic factors influencing career aspiration, career exploration, and career adaptability (*Continued*)

	Gender	Grade	Student Leader	Part-Time Jobs, Internships, or Social Practices
Amount of information	0.1101	0.1732	0.0559	0.172
Concern	0.1114	0.1705	0.0639	0.1803
Control	0.0754	0.178	0.0884	0.175
Curiosity	0.1052	0.196	0.0624	0.1822
Confidence	0.0916	0.188	0.0906	0.167
	Career Counseling	Only Child	First-Generation College Student	Family Residence
Career aspiration	0.121	0.1117	0.1081	0.1185
Career exploration	0.1169	0.1184	0.1347	0.1146
Career adaptability	0.1239	0.1029	0.1228	0.1152
Achievement aspiration	0.122	0.1072	0.1266	0.1119
Leadership aspiration	0.1032	0.1129	0.1342	0.1188
Educational aspiration	0.1211	0.1098	0.1263	0.1332
Environment exploration	0.1165	0.1056	0.1233	0.1066
Self-exploration	0.1267	0.0935	0.1345	0.1189
Purpose-systematic exploration	0.0976	0.1178	0.1388	0.1114
Amount of information	0.1184	0.1212	0.1291	0.1201
Concern	0.1151	0.1235	0.1334	0.1019
Control	0.1207	0.1087	0.1189	0.1349
Curiosity	0.1036	0.11	0.1168	0.124
Confidence	0.1264	0.0992	0.1203	0.1169

Table 5 reveals that in career aspirations and adaptability, participation in part-time jobs, internships, or social practices, grade, and career counseling experience all rank in the top three. However, participation in part-time jobs, internships, or social practices scores highest in career aspirations, while grade scores highest in career adaptability. For career exploration, grade; participation in part-time jobs, internships, or social practices; and first-generation college students rank as the top three. At the same time, the importance of different influencing factors varies across sub-dimensions. For achievement aspirations and leadership aspirations, the factors with the highest importance scores are: participation in part-time jobs, internships, or social practices, grade, and first-generation college students in the family. For educational aspirations, the factors with the highest importance scores are: participation in part-time jobs, internships, or social practices, grade, and family residence. For environment exploration, the factors with the highest importance scores are: grade, participation in part-time jobs, internships or social practices, and gender. For purpose-systematic exploration and concern, the factors with the highest importance scores are: participation in part-time jobs, internships, or social practices, grade and first-generation college students in the family. The factors with the

highest importance scores for self-exploration and amount of information are grade, participation in jobs, part-time internships or social practice, and first-generation college students in the family. The factors with the highest importance scores for control and curiosity are grade, participation in part-time jobs, internships or social practices, and family residence. The factors with the highest importance scores for confidence are grade, participation in part-time jobs, internships, or social practice, and career counseling experience.

In summary, the most influential factors are: grade, participation in part-time jobs, internships, or social practice, first-generation college students, family residence, and career counseling. As students grow older, their career knowledge and experience expand, while they face increasingly urgent challenges in future planning and career development, which significantly impacts career aspiration, career exploration, and career adaptability. Part-time jobs, internships, or social practices in career development for college students are helpful to strengthen career aspirations, stimulate career exploration, and enhance career adaptability through real-life experiences. As the first child in their family to attend colleges, students often carry higher parental expectations and a stronger sense of mission. Disparities in educational services and cultural resources between urban and rural areas significantly impact college students' career development. Career counseling, as a systematic and targeted training, guides students to set clear development goals and build professional confidence by improving their overall qualities.

4.2 Algorithm improvement analysis

By introducing feature importance weights to optimize the KNN distance metric formula, this study addresses the issue where the assumption of equal feature weights in traditional algorithms does not align with actual data characteristics. Additionally, a GA is employed to perform global optimization of model hyperparameters, thereby improving the model's predictive Acc and generalization ability. This study uses Acc, precision (P), recall (R), and weighted F1-score to evaluate the predictive performance of the machine learning model. These metrics help provide a deeper understanding of the model's performance across different dimensions. The results indicate that the improved hybrid model performs excellently in multi-dimensional predictions of student career development, providing reliable algorithmic support for related research. The formulas for each metric are as follows.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, P = \frac{\text{TP}}{\text{TP} + \text{FP}}, R = \frac{\text{TP}}{\text{TP} + \text{FN}}, F_1 = \frac{2 \times R \times P}{R + P}$$

Among them, TP denotes true positives, where both the actual classification and the model prediction are positive; TN represents true negatives, where both the actual classification and the model prediction are negative; FP indicates false positives, where the actual classification is negative but the model predicts positive; FN signifies false negatives, where the actual classification is positive but the model predicts negative. Acc represents the proportion of correctly predicted samples to the total number of samples, reflecting the overall Acc of the model's predictions. Weighted *P* calculates the proportion of samples predicted by the model as belonging to a certain category based on the sample size of each category, eliminating evaluation bias caused by category imbalance. Weighted recall rate (*R*) calculates the proportion of samples belonging to a certain category that are correctly predicted by the model, based on the sample size of each category. The weighted

F1 score is the harmonic mean of weighted P and weighted R , which comprehensively reflects the Acc and R ability of the model. Table 6 compares the use of the KNN improvement algorithm, revealing that the enhanced model demonstrates improved Acc. Table 7 evaluates the predictive performance of demographic variables across dimensions such as career aspiration, career exploration, and career adaptability.

Table 6. Comparison of random forest algorithm Acc before and after improvement

	Acc of the Random Forest Algorithm	Acc of KNN Random Forest Algorithm	Difference
Career aspiration	0.6000	0.6235	0.0235
Career exploration	0.6000	0.6235	0.0235
Career adaptability	0.6941	0.7294	0.0353
Achievement aspiration	0.5647	0.5765	0.0118
Leadership aspiration	0.5294	0.6235	0.0941
Educational aspiration	0.5765	0.6471	0.0706
Environment exploration	0.5412	0.6118	0.0706
Self-exploration	0.5412	0.5647	0.0235
Purpose-systematic exploration	0.6706	0.6471	-0.0235
Amount of information	0.6353	0.6000	-0.0353
Concern	0.6588	0.6941	0.0353
Control	0.6000	0.6471	0.0471
Curiosity	0.7176	0.7059	-0.0117
Confidence	0.6118	0.6471	0.0353

Table 7. KNN random forest algorithm prediction model evaluation

	Category	Weighted Acc	Weighted R	Weighted F1 Score
Career aspiration	Top 25%	0.62	0.78	0.69
	25%–50%	0.86	0.50	0.63
	50%–75%	0.61	0.52	0.56
	75%–100%	0.50	0.71	0.59
Career exploration	Top 25%	0.68	0.59	0.63
	25%–50%	0.57	0.67	0.62
	50%–75%	0.52	0.67	0.59
	75%–100%	0.73	0.59	0.65
Career adaptability	Top 25%	0.68	0.71	0.69
	25%–50%	0.68	0.74	0.71
	50%–75%	0.86	0.78	0.82
	75%–100%	0.71	0.67	0.69

(Continued)

Table 7. KNN random forest algorithm prediction model evaluation (Continued)

	Category	Weighted Acc	Weighted R	Weighted F1 Score
Achievement aspiration	Top 25%	0.54	0.67	0.60
	25%–50%	0.74	0.48	0.58
	50%–75%	0.60	0.47	0.53
	75%–100%	0.48	0.75	0.59
Leadership aspiration	Top 25%	0.53	0.77	0.63
	25%–50%	0.77	0.59	0.67
	50%–75%	0.53	0.60	0.56
	75%–100%	0.71	0.53	0.61
Educational aspiration	Top 25%	0.67	0.96	0.79
	25%–50%	0.76	0.57	0.65
	50%–75%	0.44	0.44	0.44
	75%–100%	0.67	0.56	0.61
Environment exploration	Top 25%	0.68	0.62	0.65
	25%–50%	0.57	0.76	0.65
	50%–75%	0.58	0.61	0.60
	75%–100%	0.62	0.48	0.54
Self-exploration	Top 25%	0.60	0.52	0.56
	25%–50%	0.43	0.47	0.45
	50%–75%	0.60	0.67	0.63
	75%–100%	0.62	0.60	0.61
Purpose-systematic exploration	Top 25%	0.65	0.83	0.73
	25%–50%	0.75	0.75	0.75
	50%–75%	0.46	0.60	0.52
	75%–100%	0.81	0.48	0.60
Amount of information	Top 25%	0.60	0.75	0.67
	25%–50%	0.61	0.65	0.63
	50%–75%	0.52	0.55	0.54
	75%–100%	0.67	0.50	0.57
Concern	Top 25%	0.82	0.78	0.80
	25%–50%	0.57	0.60	0.59
	50%–75%	0.76	0.64	0.70
	75%–100%	0.62	0.76	0.68
Control	Top 25%	0.63	0.50	0.56
	25%–50%	0.64	0.76	0.70
	50%–75%	0.68	0.65	0.67
	75%–100%	0.62	0.71	0.67

(Continued)

Table 7. KNN random forest algorithm prediction model evaluation (*Continued*)

	Category	Weighted Acc	Weighted R	Weighted F1 Score
Curiosity	Top 25%	0.62	0.72	0.67
	25%–50%	0.80	0.64	0.71
	50%–75%	0.74	0.74	0.74
	75%–100%	0.69	0.75	0.72
Confidence	Top 25%	0.74	0.61	0.67
	25%–50%	0.64	0.67	0.65
	50%–75%	0.71	0.65	0.68
	75%–100%	0.47	0.69	0.56

5 RECOMMENDATIONS

The main recommendations of this study are as follows.

1. The improved KNN Random Forest algorithm proposed in this study retains the advantages of traditional Random Forest algorithms, such as ensemble learning, resistance to overfitting, and adaptability to complex data. Through structural optimization and key parameter introduction, it achieves further Acc improvements over conventional Random Forest algorithms. The algorithm can more accurately and efficiently capture the intrinsic relationships among various influencing factors, clarify the intensity and mechanisms of different factors' impact on the prediction target, and, to some extent, addresses the issues of significant prediction bias and imprecise feature recognition in traditional algorithms when processing partial data. Notably, the introduction of GA and hyperparameter analysis enables the evaluation of model performance corresponding to hyperparameter combinations through objective functions, thereby facilitating the screening of high-quality individuals and advancing evolutionary iterations. This provides a reference for solving similar problems.
2. This study demonstrates the significant impact of part-time jobs, internships, and social practices on career development among vocational college students. Vocational colleges, which are dedicated to cultivating highly skilled professionals who serve on the front lines of production and life, particularly require students to gain rich professional experience and hands-on practice. Such experiences help students establish correct career perspectives, enhance professional competencies, and accumulate relevant experience in occupational environments. The research also reveals that part-time jobs, internships, and social practices significantly influence students' career aspiration, exploration, and adaptability. Therefore, colleges should further emphasize and strengthen guidance and management of students' part-time work internships and social practices. Firstly, colleges should establish comprehensive systems to differentiate management priorities between part-time jobs and professional internships, developing corresponding management methods. For college-organized internships, clear implementation procedures, safety requirements, disciplinary norms, and evaluation standards must be defined. Secondly, colleges should form specialized teams to provide multifaceted guidance. For example, conducting training sessions on career planning, rights protection, and time management for part-time activities

can help students balance work and studies while avoiding pitfalls. For internships, position-specific skill guidance should be integrated with professional characteristics, with regular progress tracking and timely resolution of professional challenges and psychological issues encountered during internships. Finally, colleges should strengthen collaboration with enterprises and communities to establish stable internship bases and standardize practice content and evaluation criteria to ensure quality. At the same time, colleges should open up the communication channels, strengthen the linkage with students, parents, and enterprises, grasp the students' practice in time, and coordinate to solve the contradictions and problems in practice.

3. This study examines the significant impact of career counseling on the development of vocational college students. Career counseling ranks among the top three variables in terms of importance in career aspiration and career adaptability. However, compared to part-time jobs, internships, and social practices, the influence of career counseling remains relatively limited, necessitating further enhancement to maximize its positive effects. Career counseling plays a crucial supporting role in student growth. To strengthen career counseling for college students, the first step is to establish a comprehensive, multi-level counseling system that meets students' developmental needs throughout their academic journey. Incorporating career counseling into talent cultivation plans, refining curriculum systems by grade, and implementing targeted teaching approaches are essential. For lower grades, focus should be placed on career planning and occupational overview, while for upper grades, emphasize skill enhancement and decision-making guidance, integrating practical content such as job-seeking techniques and career assessments. Secondly, professional training should be combined with practical activities, creating diverse practice platforms through events like career planning competitions, alumni sharing sessions, and corporate study tours, enabling students to deepen their understanding of careers through hands-on experience. Finally, improving collaborative mechanisms and personalized services, providing one-on-one counseling tailored to different majors and needs, with special attention to students facing unique challenges, will effectively enhance counseling outcomes and support students' career development.

6 REFERENCES

- [1] N. Papé, I. Mahmud, R. Hasan, and N. Uddin, "Learning needs including preferences and digital technologies: A study of mature students in higher education in England," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 14, pp. 7–19, 2025. <https://doi.org/10.3991/ijim.v19i14.56921>
- [2] P. Yang, X. Wang, J. Zhang, S. K. Sivaraman, and P. Song, "Exploring the impact of artificial intelligence on university students' perception of slow employment: A psychological and behavioral analysis," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 18, pp. 92–105, 2025. <https://doi.org/10.3991/ijim.v19i18.57243>
- [3] E. A. York, "Gender differences in the college and career aspirations of high school valedictorians," *Journal of Advanced Academics*, vol. 19, no. 4, pp. 578–600, 2008. <https://doi.org/10.4219/jaa-2008-830>
- [4] F. Nadeem and R. Khalid, "The relationship of gender role attitudes with career aspirations and career choices among young adults," *Pakistan Journal of Psychological Research*, vol. 33, no. 2, pp. 455–471, 2018.

- [5] T. S. Chang, H. C. Wang, and S. T. Chen, "A longitudinal study of educational aspirations and occupational aspirations among college students," *Journal of Research in Education Sciences*, vol. 67, no. 3, pp. 147–176, 2022.
- [6] M. A. Al-Bahrani, S. M. Allawati, Y. A. Abu Shindi, and B. S. Bakkar, "Career aspiration and related contextual variables," *International Journal of Adolescence and Youth*, vol. 25, no. 1, pp. 703–711, 2020. <https://doi.org/10.1080/02673843.2020.1730201>
- [7] J. Zikic and U. C. Klehe, "Job loss as a blessing in disguise: The role of career exploration and career planning in predicting reemployment quality," *Journal of Vocational Behavior*, vol. 69, no. 3, pp. 391–409, 2006. <https://doi.org/10.1016/j.jvb.2006.05.007>
- [8] H. Chen *et al.*, "Career exploration of high school students: Status quo, challenges, and coping model," *Frontiers in Psychology*, vol. 12, p. 672303, 2021. <https://doi.org/10.3389/fpsyg.2021.672303>
- [9] S. A. Stumpf and M. C. Lockhart, "Career exploration: Work–role salience, work preferences, beliefs, and behavior," *Journal of Vocational Behavior*, vol. 30, no. 3, pp. 258–269, 1987. [https://doi.org/10.1016/0001-8791\(87\)90004-2](https://doi.org/10.1016/0001-8791(87)90004-2)
- [10] S. Zhang and L. Ye, "A study of the development characteristics of juvenile career exploration," *Journal of Shenyang Normal University (Social Science Edition)*, vol. 34, no. 5, pp. 83–86, 2010.
- [11] D. N. Wang *et al.*, "Revision of student career construction inventory," *Chinese Journal of Clinical Psychology*, vol. 24, no. 1, pp. 100–128, 2016.
- [12] B. Kracke, "Parental behaviors and adolescents' career exploration," *Career Development Quarterly*, vol. 45, no. 4, pp. 341–350, 1997. <https://doi.org/10.1002/j.2161-0045.1997.tb00538.x>
- [13] M. D. C. Taveira, M. C. Silva, M. L. Rodriguez, and J. Maia, "Individual characteristics and career exploration in adolescence," *British Journal of Guidance and Counselling*, vol. 26, no. 1, pp. 89–104, 1998. <https://doi.org/10.1080/03069889808253841>
- [14] M. L. Savickas and E. J. Porfeli, "Career adapt–abilities scale: Construction, reliability, and measurement equivalence across 13 countries," *Journal of Vocational Behavior*, vol. 80, no. 3, pp. 661–673, 2012. <https://doi.org/10.1016/j.jvb.2012.01.011>
- [15] X. Y. Zhao and C. Guo, "Review and commentary on foreign studies of career adaptability," *Advances in Psychological Science*, vol. 18, no. 9, pp. 1503–1510, 2010.
- [16] A. Hirschi, "Career adaptability development in adolescence: Multiple predictors and effect on sense of power and life satisfaction," *Journal of Vocational Behavior*, vol. 74, no. 2, pp. 145–155, 2009. <https://doi.org/10.1016/j.jvb.2009.01.002>
- [17] O. A. Olugbade, "The career adapt–abilities scale–Nigeria form: Psychometric properties and construct validity," *Journal of Vocational Behavior*, vol. 95, pp. 111–114, 2016. <https://doi.org/10.1016/j.jvb.2016.08.006>
- [18] X. Xiao, "Research on career adaptability of students of agricultural and forestry university from the perspective of perceived social support–Based on the empirical study of five colleges," *China Forestry Economy*, no. 6, pp. 33–38, 2023.
- [19] H. Zulfiani and N. M. Khaerani, "Interrelation between career adaptability and family support, gender and school type," *Jurnal Psikologi Integratif*, vol. 8, no. 2, pp. 80–91, 2021. <https://doi.org/10.14421/jpsi.v8i2.1888>
- [20] Z. Liu and J. Zhang, "The dynamics of social and psychological capital conversion in career adaptability: A network analysis," *Journal of Psychology in Africa*, vol. 35, no. 5, pp. 575–586, 2025. <https://doi.org/10.32604/jpa.2025.067476>
- [21] C. C. Kiu, "Data mining analysis on student's academic performance through exploration of student's background and social activities," in *Proceedings of the 4th International Conference on Advances in Computing, Communication and Automation (ICACCA)*, Subang Jaya, Malaysia, 2018, pp. 1–5. <https://doi.org/10.1109/ICACCA.2018.8776809>

- [22] M. Pellagatti, C. Masci, F. Ieva, and A. M. Paganoni, "Generalized mixed-effects random forest: A flexible approach to predict university student dropout," *Statistical Analysis and Data Mining*, vol. 14, no. 3, pp. 241–257, 2021. <https://doi.org/10.1002/sam.11505>
- [23] M. A. Gregor and K. M. O'Brien, "Understanding career aspirations among young women: Improving instrumentation," *Journal of Career Assessment*, vol. 24, no. 3, pp. 559–572, 2016. <https://doi.org/10.1177/1069072715599537>
- [24] C. Xu, "Career decision-making self efficacy and it's relation to anxiety and career exploration," M.S. thesis, Central China Normal University, Wuhan, China, 2008.
- [25] Z. J. Hou, S. A. Leung, X. Li, X. Li, and H. Xu, "Career adapt-abilities scale—China form: Construction and initial validation," *Journal of Vocational Behavior*, vol. 80, no. 3, pp. 686–691, 2012. <https://doi.org/10.1016/j.jvb.2012.01.006>

7 AUTHORS

Fei Gao is a researcher at the School of Public Affairs, Zhejiang Shuren University, and holds a Doctor of Education degree. Her research primarily focuses on higher vocational education (E-mail: gaofei@zjsru.edu.cn).

Jing Yang is an associate researcher at Hebei Women and Children Activity Center in Shijiazhuang. She holds a Master of Education degree, and her research interests lie in educational administration (E-mail: pljinger@126.com).