International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-15

https://doi.org/10.70917/ijcisim-2025-0210
Article

Research on the Professional Development Path of Preschool Teachers under the Analysis of Intelligent Educational Big Data

Jing Chen *

Lishui University, Lishui, Zhejiang, 323000, China; cj96758@163.com

Abstract: Most of the current research on teachers' professional development is based on literature and theory, lacking empirical support. In this regard, a study on the professional development path of preschool teachers under the support of intelligent education big data is carried out. 500 young teachers in colleges and universities belonging to the first-level discipline of "preschool education" and under the age of 40 were selected as data collection targets, and data collection and pre-processing were carried out. On this basis, the clustering algorithm and IF-IDF algorithm were used to construct a teacher portrait model based on data mining technology, and the model was applied to preschool education in colleges and universities, thus designing a professional development path for preschool teachers that integrates the teacher portrait model, and then carrying out an empirical analysis of the path. After the intervention of the teaching experiment, it was found that the differences between the experimental group and the control group in teaching skills, theoretical knowledge of teaching and values were at a significant level, $P<0.05$, indicating that the integration of the teacher portrait model in the professional development path of preschool teachers is more conducive to the professional development of preschool teachers.

Keywords: clustering algorithm; IF-IDF algorithm; preschool education; teacher professional development paths

1. Introduction

In response to the aging, the population fertility rate continues to decline, China's three-child policy officially landed, preschool education once again into people's view. It is true that only by solving the problem of expensive and difficult pre-school education can the incentives of the population policy be truly realized, and can help the contemporary young people to remove the worries of childbearing, which is especially important for young people of the right age [1]. At the same time, the pre-school period is also an important period for children's rapid cognitive development and the evolution of all aspects of the ability [2]. If quality preschool education is provided in a timely manner, it can not only provide sufficient support for the healthy physical and mental development of young children, but also lay a solid foundation for future learning after entering school. A teacher group with excellent professionalism and reasonable team structure can provide support for pre-school children to better receive knowledge education, ideological and moral education [3-5]. However, it is regrettable that under the combined effect of many factors such as social factors, economic factors, cultural factors, etc., together with the lack of attention from all sectors of the society and relevant departments, it leads to the fact that there is still a certain amount of room for improvement and potential for perfection in the construction of preschool teachers' teams in most regions [6]. The arrival of the big data era has broken through the time and space limitations of preschool education and teaching, promoted the in-depth integration of digital science and technology and education, and opened up a new journey for the digital development of education [7]. All this is becoming a new kinetic energy for digital talent training, and at the same time, it puts forward new requirements for the professional development of preschool teachers, which requires teachers to rely on the opportunity of digital construction, keep up with the new requirements of education digitization, open up a new track for education development, promote the reform of digital teaching, and promote their own growth and development [8-9]. Therefore, it is particularly important to clarify the path of professional development of preschool teachers, and then deeply understand the impact of the big data era on the professional development of teachers.

Teachers' professional development competencies cover many aspects of knowledge, skills, literacy, and qualities necessary for professional development [10]. In the era of big data, data has become a new production factor driving the development of education, and the role played by the dynamics of its professional development is of great significance for the creation of a specialized and innovative teaching force [11-12]. Gafarov and Galimyanov [13] explored the necessity of using big data methods in educational activities, and clearly pointed out that big data and machine learning methods are the key factors in determining the professional development of teachers, standing on top of the technology, flexible selection and use of technology, so that the technology is perfectly integrated with education and teaching. Leskina [14] proposes a program model for managing teacher effectiveness using big data, and also introduces the concept of “complex professional competencies” for teachers working with big data, which provides a methodological approach to understanding teacher professional development. This provides a methodological approach to understanding teacher professional development. Besbes et al. [15] explored the use of big data analytics in education by tracking the performance of 1000 students and 120 teachers to facilitate teacher professional development and improve learning outcomes. Cui and Zhang [16] pointed out that the professional development path of teachers in the era of big data includes, cultivating the awareness of big data, enhancing the wisdom of big data, establishing a new model of cooperative alliance, establishing the concept of lifelong learning, and continuously updating the knowledge structure. Sun et al. [17] argued that the online professional development platform has accumulated a large amount of behavioral data about teachers, and these data also provide an important opportunity for the assessment of teachers' professional development, and by analyzing and mining these data through big data technology, the can strengthen the foundation of teachers' professional development and promote the improvement of their own teaching ability. The above literature has explored the mechanism of big data technology on teachers' professional development, which provides a reference for the research, however, there are still relatively few studies on the professional development of preschool teachers.

In this paper, on the basis of determining the object, type and source of this data collection, data center sharing and network crawling are used to collect preschool teacher portrait data, and further data preprocessing is required to ensure data availability. Within the scope of data mining technology, the teacher portrait model is constructed with the help of clustering algorithm and IF-IDF algorithm. According to the intelligent diagnosis and teacher behavior analysis features of the teacher portrait model, a professional development path for preschool teachers integrating the teacher portrait model is designed, and the path is validated and analyzed using scales and statistical analysis.

2. Teacher Profiling Model Based on Data Mining Techniques

Teacher portraits also provide data support for school administrators to make relevant decisions, through the analysis of teacher group portraits and statistical analysis of relevant labels can enable school administrators to clearly grasp the overall professional development of teachers, and timely intervention to promote teacher professional development. This study takes the professional development of preschool teachers in college A as an example, analyzes the professional development of teachers in the college based on their portraits, and identifies the problems of teacher development, so as to formulate corresponding strategies to cope with them. The contents of the teacher portrait model construction based on data mining technology are as follows:

2.1. Multi-source data acquisition and pre-processing

2.1.1. Data Acquisition Objects

Due to the competitive mechanism of quantitative assessment implemented by colleges and universities in the professional development of teachers, there is a lack of effective interaction and communication among young teachers, and most of the teachers rely on their daily social situation and intuitive feelings to choose the object of scientific research cooperation or exchange, which is relatively limited in scope and prone to produce information barriers [18]. In addition, in recent years, colleges and universities have introduced the “promotion or departure” selection system, and young teachers are subject to the strong incentive effect brought by the elimination mechanism, and the value of professional development has become more and more prominent. Therefore, this paper selects 500 young teachers who belong to the first-level discipline “preschool education” and are under 40 years old as the data collection object of teacher portrait.

2.1.2. Types and sources of data

Rich and sufficient data is the foundation for the success of constructing the portrait of teachers' professional development. Before determining the data sources, the kinds of data needed in this paper in constructing the portrait of teachers' professional development should be analyzed. According to the existing research results, the data needed to construct the portrait are divided into the following three categories, which are teachers' basic attributes, teachers' teaching attributes, and teachers' research attributes. After determining the types of data, according to the level of informationization construction and the feasibility of network crawler in college A, so as to determine the source of data, there are two main sources of data, one is from the various systems of the school, and the other is from the open Internet environment.

2.1.3. Data collection methods

For different data sources, the way of data collection is different, there are two main methods of data collection in this paper, which are data center sharing method and web crawler method.

(1) Data Center Sharing Approach

The personnel office system of colleges and universities stores the demographic characteristics information of teachers, and the teaching management system and the library academic achievement repository store a large amount of professional development data of teachers. And these systems are generally independent of each other and the data are not correlated, so when collecting data, it is often necessary to extract relevant data from the database of each system. As the level of digital construction in colleges and universities continues to improve, in order to better manage data and achieve resource sharing, many colleges and universities have established data center platforms to efficiently integrate various business systems together. A data center platform in colleges and universities is able to achieve the collection, storage, management and sharing of data, so this paper collects data from multiple systems together through the school's data center platform. In order to ensure data security, the collection of multi-source data is realized based on the API interface on the school data center platform. When calling the API interface service, the application identification is passed in and the parameters are signed with an algorithm key to complete the identity authentication. When the data center platform receives a legitimate request, it will use the same algorithm to sign and compare whether the two signatures before and after are the same, so as to verify whether data tampering has occurred during the transmission of the protocol, and if they are not the same, then the API service will be rejected to avoid data leakage. In order to ensure the timeliness of the API, the validity time of 30 seconds is also set in the request header, and requests exceeding the timeliness will be regarded as illegitimate requests, and will also be refused to be called by the service.

(2) Web crawler approach

Design the crawler approach in this paper: firstly, input the relevant data in the target URL address for retrieval, simulate sending a network request, get the Response response content, and find the corresponding web page HTML content in it [19]. If there is target data in the HTML content, the corresponding text is parsed directly. If not, send a request to the sub-URL address of the HTML of the web page, look for the target data in the response content, and carry out the search in turn until the target data is found. Then parse the target data to get the text content, and finally save the data in the specified CSV file for subsequent processing.

2.1.4. Data pre-processing

In order to ensure the completeness, uniformity, and accuracy of the dataset, data preprocessing work is needed before using the data, so that it can be transformed into data that can be used to construct a research portrait of university teachers [20-21]. The process of data preprocessing varies for different dataset attributes and tasks, and generally includes three steps: data cleaning, data conversion, and data integration.

(1) Data cleansing

Data cleansing refers to screening, filling, deleting and other operations on data that are duplicated, redundant, missing and abnormal to ensure the consistency and integrity of the data, and to effectively ensure the accuracy of the data and the quality of decision-making.

(2) Data Conversion

Data conversion aims to convert raw data into a format and structure suitable for a particular analytical task, where data in many different formats are converted into a uniform format. This makes the data easy to use and manage, and also improves the processing efficiency and accuracy of the data.

(3) Data Integration

Data integration is the final step of the whole data preprocessing, aiming at integrating data from different data sources into a complete data set for subsequent data analysis and mining.

2.2. Model Construction

In order to make the teacher portrait more concrete and visual, teachers cannot be depicted from only one perspective. In this model, the teacher's basic attribute tag model is represented by UserInfo, the teacher's teaching attribute tag model is represented by TeAttr, and the teacher's research attribute tag model is represented by AcAttr. UserInfo, TeAttr, and AcAttr are composed of sub-models ThemeTopic and Tag of different dimensions, respectively.

2.2.1. Teacher Basic Attribute Labeling Model

Teacher basic attribute labeling model is shown in Fig. 1, Teacher basic attribute labeling model is represented by UserInfo, which contains two topics, i.e., Topic=<Basic Information, Job Information>. Where basic information is the teacher's personal profile information that can be used to identify the teacher's identity, BasicInfo=<Name, Gender, Age, Education>. Job information is the information related to the teacher's job, job information=<work number, title, faculty>. All the tags in the teacher's basic attribute tag except age can be obtained by directly transforming the information collected from the teaching management system system.

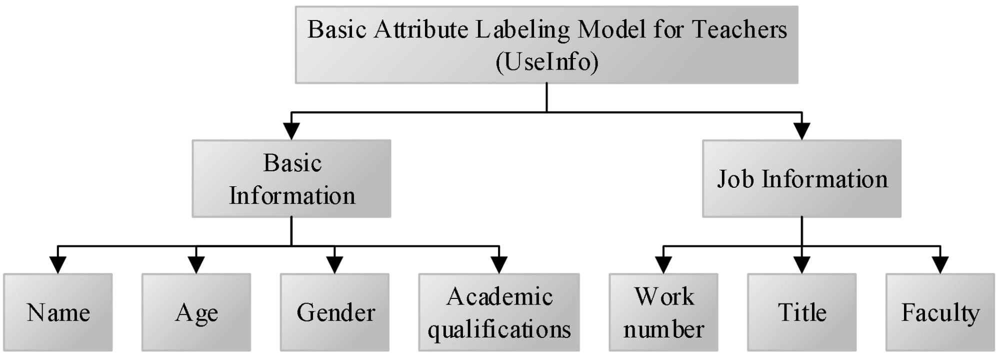

Figure 1. Teacher basic attribute label model.

2.2.2. Teacher Instructional Attribute Labeling Model

The teacher teaching attribute labeling model is shown in Figure 2, and the teacher teaching attribute labeling model is represented by TeAttr, which contains three topics, i.e., Topic=<Teaching Performance, Teaching Measurement, and Teaching Achievement>. Where teaching performance refers to the results of the school's assessment of teachers' teaching performance. Teaching assessment refers to the results of students' and colleges' evaluation of teachers' teaching work. Teaching Achievement refers to the results of teachers' teaching work, including teachers' participation in teaching construction and research, winning teaching achievement awards and teaching quality awards, and guiding students to participate in competitions and win prizes. Teaching Achievement = <Teaching Construction, Teaching Award, Competition Guidance>.

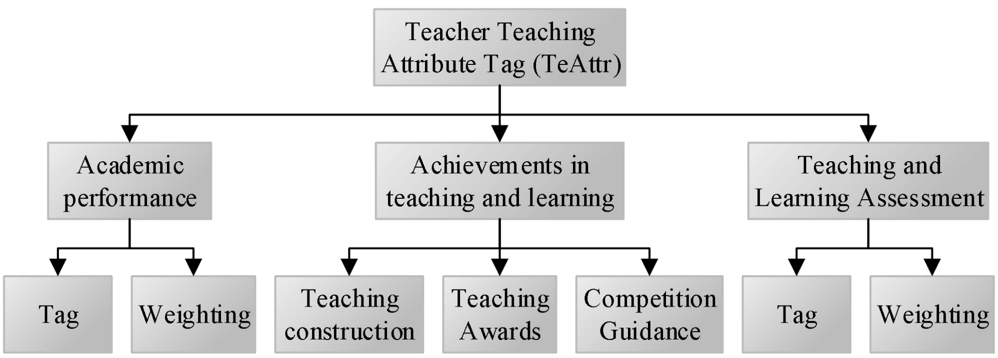

Figure 2. Teacher teaching attribute label model.

2.2.3. Faculty Research Attribute Labeling Model

The teacher research attribute tag model is represented by AcAttr, which also contains three topics, Topic=<Research Achievement, Research Performance, Research Direction>. Among them, research achievement refers to the research results achieved by teachers in research, including research projects, publications, research papers and research awards. Research performance refers to the assessment results of the university on teachers' scientific research work, and research direction indicates the direction of teachers' research themes and teachers' academic influence, and the model of teachers' scientific research attribute labels is shown in Figure 3.

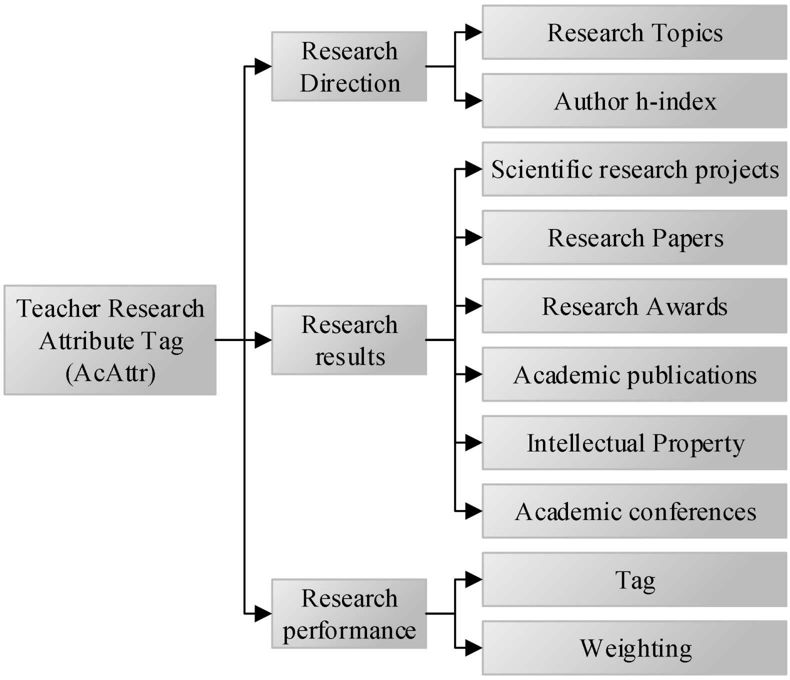

Figure 3. Teacher research attribute label model.

In this study, the IF-IDF algorithm is used for keyword extraction, the higher the number of times a word appears in a document, the higher its importance, but its importance decreases as the frequency of its appearance in the corpus grows.

Word Frequency (TF) is the frequency of occurrence of a word in a given document and is a reflection of the importance of the word in the document. Assuming that a word in a given document is defined as t, its importance can be expressed as:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

In the formula, the numerator represents the number of times the word occurs in the file d_j and the denominator represents the sum of all the occurrences of the word in the file d_j.

Inverse Document Frequency (IDF) is used to measure the universal importance of a word. The IDF of a particular word is obtained by dividing the total number of files by the number of files containing the word, and the result is logarithmic with a base of 10, as follows:

$$idf_i = \lg \frac{|D|}{\{j : t_i \in d_j\}} \tag{2}$$

where $|D|$ denotes the sum of the number of all files in the corpus. $\left|\{j : t_i \in d_j\}\right|$ denotes the number of files containing the word t_i. If the word is not in the profile, the denominator will be equal to zero, so in general $1+\left|\{j : t_i \in d_j\}\right|$ is used. multiplying Eqs. 3-1 and 3-2 gives the TF-IDF value as shown in

Eq. (3). For:

$$tfidf_{i,j} = tf_{i,j} \times idf_i \tag{3}$$

By calculating and comparing the TF-IDF value of each word in the document, the three keywords with the largest IF-IDF were selected as the labels for that teacher's professional development. For the keywords in teachers' professional development, because there are so many types of outcomes involved, some simplification was done in this study.

2.2.4. Label weight determination

The teacher portrait model designed in this study is an annual portrait, and considering the need to make a long-term portrait of teachers, this study assigns weights to labels in two dimensions: frequency and time.

(1) Frequency-based label weighting

Assuming that a teacher generates a certain label in any given year, the number of times the label appears increases by one. It is concluded that the more times a label appears, the more it reflects the teacher's professional development level. Therefore, define the weight of label t_i for teacher u as follows:

$$P_{i,u} = \frac{Freq_{i,u}}{\text{Time}} \tag{4}$$

where $Freq_{t_i,u}$ denotes the number of times teacher u's label t appears, and Time denotes the time (in years) that the portrait represents. That is, assuming that teacher A has three "first paper" labels in five years, the frequency weighting for that label is P (first paper) = 0.6.

(2) Time-Based Label Weighting

The core of time dimension based weighting is that labels generated by the teacher recently should have a larger weight, and the further back in time they were generated, the smaller their weight. Therefore, the weight of a single occurrence of label t_i is defined as follows:

$$TW_{t,n} = \frac{T_{1,n}}{\text{Time}} \tag{5}$$

where n denotes that this is the first occurrence of the label, $T_{t_i,n}$ denotes the time of the n-th occurrence of label t_i, and Time denotes the time (in years) indicated by the portrait.

Define the time dimension based weights of label t_i on teacher u as follows:

$$TW_{t_i,u} = \sum_{n=1}^{n} TW_{t_i,n} \tag{6}$$

i.e., by adding the weights calculated for each occurrence of the tag, the total weight of the tag based on the time dimension is obtained, and the time weighting of the tag is $0.2+0.6+0.8=1.6$.

(3) Determination of total label weights

After determining the weights of the labels based on the frequency dimension and the time-based dimension, the total weights of the labels need to be determined. In this study, frequency and time are assigned a weight of 0.6 and 0.4, respectively, i.e., the total weight of label t_i on teacher u is defined as follows:

$$W_{t_i,u} = 0.6P_{t_i,u} \times 0.4TW_{t_i,u} \tag{7}$$

2.2.5. Clustering algorithm

In order to differentiate teachers and study the characteristics of different groups of teachers, this study adopts the K-means clustering method in data mining technology to cluster teachers according to preschool professional development labels. The K-means clustering algorithm generally uses the Euclidean distance as a measure of the similarity between data objects, and the degree of similarity is inversely proportional to the distance between the data objects, and the greater the degree of similarity, the smaller the distance. The K- means clustering algorithm is an unsupervised clustering algorithm, before the clustering begins, it is necessary to pre-set the initial number of clusters K and specify the K initial clustering centers, and then divide the clustered samples into K clusters, calculate the similarity between the clustered samples and the clustering centers within each cluster, and continuously iteratively update the position of the clustering centers in order to make all clustered objects within the clusters have a high degree of similarity until the cluster has the is the highest and the sum of squared errors is the smallest, and the clustering is stopped to produce results.

The basic flow of data processing of K-means cluster analysis algorithm. Its specific realization steps are as follows:

(1) Randomly select K objects from the sample data as the center of the initial clustering.

(2) Calculate the Euclidean distance from each sample to each cluster center, and assign the sample to the cluster center category with the closest distance.

(3) After all samples are assigned, recalculate the centers of the K clusters.

(4) Compare with the K clustering centers obtained from the previous calculation, if the clustering center changes, recalculate the Euclidean distance between each sample data and the new clustering center, and reallocate the samples to the category belonging to the new clustering center with the closest distance; if the clustering center doesn't change, stop clustering and output the results.

K-means clustering is to minimize the sample and prime squared error as the objective function, the cluster center of each cluster and the cluster of sample points within the squared distance error sum is called the degree of aberration, for a clustering cluster, the lower its degree of aberration, on behalf of the cluster of samples between the points of the smaller distance, the closer the connection; aberration degree, on behalf of the cluster of samples between the points of the larger distance, the looser the structure. Although the degree of distortion will decrease with the increase of the clustering category, but for a certain degree of differentiation of the samples, the degree of distortion will be greatly reduced when a critical point is reached, and then the decline tends to level off, this critical point can be regarded as a better clustering effect of the K-value of the representative point. The contour coefficient is also an important indicator for evaluating the degree of density and dispersion of clustering, and its calculation formula is as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{8}$$

where $a(i)$ represents the degree of cohesion of the sample points, which is calculated as follows:

$$a(i) = \frac{1}{n-1}\sum_{j \neq i}^{n} distance(i, j) \tag{9}$$

where j represents the other sample points within the same class as sample i, and $distance$ represents the distance from i to j that is sought. So the smaller $a(i)$ indicates the closer the class is.

The calculation of $b(i)$ is similar to that of $a(i)$, the difference is that multiple values need to be obtained by traversing the other class clusters, and then selecting the smallest value from these values as the final result.

From the above equation, it can be found that when $a(i) < b(i)$, i.e., when the intra-class distance is smaller than the inter-class distance, then the clustering results are better. At this time, the value of S will converge to 1. The closer the value of S converges to 1, it means the more obvious its contour. On the contrary, when $a(i) > b(i)$, the intra-class distance is greater than the inter-class distance, the clustering results are loose. At this time the value of S will tend to -1, the more the value of S tends to -1 then its profile is less significant, the worse the clustering results. From this, it can be obtained that the range of values of S is $[-1, 1]$, the larger the contour coefficient, the better the clustering results. Based on the results of the clustering analysis of the professional development of college teachers, appropriate strategies can be developed to deal with it.

2.3. Results and analysis

2.3.1. Tag Weighting Analysis

(1) Validation analysis of TF-IDF keyword extraction algorithm

(a) Evaluation indexes

On the basis of the research data in this paper, the performance of the TF-IDF keyword extraction algorithm in the teacher portrait model used in this study needs to be evaluated, and in this study, the precision rate P, the recall rate R, and the F-score are selected as the performance evaluation criteria of the keyword extraction algorithm.

The keywords of category A are the correct keywords that were extracted, the keywords of category B are the correct keywords that were not extracted, the keywords of category C are the keywords that were extracted but incorrect, and the keywords of category D are the keywords that were not extracted and incorrect.

The precision rate is calculated by the formula:

$$P = \frac{A}{A+C} \tag{10}$$

The formula for calculating the recall rate:

$$R = \frac{A}{A+B} \tag{11}$$

The F-score is the inverse average of precision and recall, which allows for a more integrated and comprehensive evaluation of performance by taking both precision and recall into account. To wit:

$$F1 = \frac{2*P*R}{P+R} \tag{12}$$

(b) Performance Comparison Experiments with Other Algorithms

The TF-IDF algorithm is experimented and performance comparison with four algorithms, namely BM (Boolean Model), STF (Summation of Word Frequency), PositionRank (Positional Sorting), and TextRank (Text Sorting), is conducted on the research data of this paper.The performance comparison between the TF-IDF algorithm and the other algorithms (extracting 10 keywords) is shown in Table 1, the performance comparison of TF-IDF algorithm with other algorithms (20 keywords extracted) is shown in Table 2, and the performance comparison of TF-IDF algorithm with other algorithms (30 keywords extracted) is shown in Table 3. Combining Table 1, Table 2 and Table 3, it can be seen that the TF-IDF algorithm has the best performance in the three evaluation indexes of Recall, Precision and F1-Measure, and the keyword extraction effect is more obvious compared to the other four algorithms, which ensures the rigor of the results of the subsequent research.

Table 1. Performance comparison of algorithms (Extract 10 key words).

Algorithm	R	P	F1
TF-IDF	0.869	0.866	0.867
BM	0.850	0.763	0.804
STF	0.727	0.709	0.718
PositionRank	0.716	0.692	0.704
TextRank	0.601	0.670	0.634

Table 2. Performance comparison of algorithms (Extract 20 key words).

Algorithm	R	P	F1
TF-IDF	0.859	0.839	0.849
BM	0.732	0.685	0.708
STF	0.696	0.683	0.689
PositionRank	0.624	0.658	0.641
TextRank	0.621	0.608	0.614

Table 3. Performance comparison of algorithms (Extract 30 key words).

Algorithm	R	P	F1
TF-IDF	0.896	0.836	0.865
BM	0.758	0.809	0.783
STF	0.703	0.761	0.731
PositionRank	0.695	0.707	0.701
TextRank	0.618	0.628	0.623

(2) Keyword Extraction for Professional Development of Preschool Teachers

After verifying the TF-IDF keyword extraction algorithm in the teacher portrait model, we carried out the keyword extraction for professional development of preschool teachers using the TF-IDF keyword extraction algorithm on the above research data, and the keyword extraction results are shown in Table 4.

Based on the data in the table, it can be seen that the TF-IDF algorithm extracted 22 keywords for preschool teachers' professional development, and the top three keywords in terms of TF-IDF value were used as the preschool teachers' professional development labels, which were teachers' basic attributes, teachers' teaching attributes, and teachers' scientific research attributes, and the corresponding TF-IDF values were 0.0843, 0.0863, and 0.0896, which were convenient for the subsequent label weight analysis work.

Table 4. Keyword extraction results.

Key words for the professional development of preschool education teachers	Number of files	Word frequency	TF	IDF	TF-IDF	Rank
Name	33	28	0.0115	1.9039	0.0218	22
Age	34	52	0.0213	1.8909	0.0402	21
Gender	49	62	0.0254	1.7322	0.0439	20
Educational background	49	65	0.0266	1.7322	0.0461	19
Employee ID	53	67	0.0274	1.6981	0.0465	18
Department	60	73	0.0299	1.6443	0.0491	16
Mathematical achievements	82	83	0.0339	1.5086	0.0512	14
Teaching construction	93	85	0.0348	1.4539	0.0505	15
Teaching awards	93	87	0.0356	1.4539	0.0517	13
Competition guidance	124	87	0.0356	1.3290	0.0473	17
Research topic	128	102	0.0417	1.3152	0.0549	11
Scientific research project	128	103	0.0421	1.3152	0.0554	10
Research paper	138	104	0.0425	1.2825	0.0546	12
Scientific research award	140	124	0.0507	1.2763	0.0647	9
Academic works	143	128	0.0524	1.2671	0.0663	8
Intellectual property Rights	154	141	0.0577	1.2349	0.0712	7
Academic conference	164	146	0.0597	1.2076	0.0721	6
Scientific research achievements	173	161	0.0658	1.1844	0.0780	5
Research direction	197	176	0.0720	1.1280	0.0812	4
Basic attributes of teachers	198	183	0.0748	1.1258	0.0843	3
Teaching attributes of teachers	205	190	0.0777	1.1107	0.0863	2
The research attribute of teachers.	207	198	0.0810	1.1065	0.0896	1

(3) Analysis of label weighting results

After determining the preschool teachers' professional development labels, according to the principle of weight calculation in the model of this paper, the teachers' basic attributes, teachers' teaching attributes, and teachers' scientific research attributes are assigned, and the results of label weighting are shown in Table 5. Based on the data performance in the table, it can be seen that the weight values of teachers' basic attributes, teachers' teaching attributes, and teachers' research attributes are 0.2676, 0.3485, and 0.3839 respectively.

Table 5. Label weight result.

Professional development label for preschool Education teachers	Frequency-based label weighting	Time-based label weighting	Total weight
Basic attributes of teachers	0.2545	0.2872	0.2676
Teaching attributes of teachers	0.3631	0.3266	0.3485
The research attribute of teachers	0.3824	0.3862	0.3839

2.3.2. Analysis of clustering results

Preschool teachers' professional development labeling is a highly refined and labeled way of presenting preschool teachers' professional development data. Based on the research data and label weights in this paper, the level of professional development of preschool teachers is clustered, which helps to manage different groups of preschool teachers in a more organized way in the subsequent teaching process, and also helps teachers to improve their teaching methods. The K-means unsupervised clustering algorithm in the model of this paper clusters the labels of preschool teachers' professional development, and after completing the clustering operation, the clustering results need to be analyzed and interpreted, and then combined with the research questions and data characteristics, the different teacher class groups formed by clustering are posted. Combined with the results of the calculation of the contour coefficient to comprehensively determine the best clustering number K value. After calculation, the K value was finally determined to be 3, so it was determined that the level of preschool teachers' professionalism was divided into 3 categories. The results of the clustering of preschool teachers' professional development labels are shown in Table 6, with the smallest proportion of the number of people accounted for by cluster 3 (118, 23.6%), a smaller proportion of the number of people accounted for by cluster 2 (147, 29.4%), and the largest proportion of the number of people accounted for by cluster 1 (235, 47.0%). An in-depth analysis of the teacher educators in each of the clusters showed that Cluster 1 performed slightly worse on both the Teacher Teaching Attributes and Teacher Research Attributes labels, Cluster 2 performed less well on the Basic Teacher Attributes and Teacher Research Attributes labels, and Cluster 3 did not perform as well as the other Teacher Research Attributes labels on both the Basic Teacher Attributes and Teacher Teaching Attributes labels. Based on the characteristics of the above three clusters and the understanding of the reality, cluster 1 is categorized as "novice" preschool teachers, cluster 2 is categorized as 'skilled' preschool teachers, and cluster 3 is categorized as "expert" preschool teachers. "The model of this paper not only can directly observe the pre-school teachers, but also can help the pre-school teachers to be more skillful. The model can not only directly observe the professional level of preschool teachers, but also help preschool teachers to have a more intuitive understanding of their own professional development. In conclusion, teachers' professionalism is crucial to the development of the education field. By gaining a deeper understanding of their individual and group characteristics, as well as providing emotional support through the model in this paper, future teachers with strong professional competencies can be better trained to promote the digital transformation of education, improve the quality of teaching and learning, as well as facilitate the innovative development of education.

Table 6. Clustering results of teacher professional development labels.

Professional development label for preschool Education teachers	Clustering		
	Group 1	Group 2	Group 3
Basic attributes of teachers	0.896	0.247	0.253
Teaching attributes of teachers	0.234	0.793	0.512
The research attribute of teachers	0.225	0.288	0.942
Number of people (proportion)	235 (47%)	147 (29.4%)	118 (23.6%)

3 Exploring the development path of preschool teachers

3.1. Designing preschool teacher development pathways

Chapter 2 constructs a teacher portrait model based on data mining data, analyzes in detail the professional development level of preschool teachers, and derives three categories of teachers, which are "novice" preschool teachers, "skilled" preschool teachers, 'expert' preschool teachers, and "expert" preschool teachers. "Expert" preschool teachers. The above analysis enables school administrators to have a clear picture of the overall professional development of teachers and to intervene in a timely manner in order to promote the professional development of teachers. In conclusion, the teacher portrait model not only provides data support for school administrators to make relevant decisions, but also lays a theoretical foundation for the design of preschool teacher development paths. The design of preschool teacher development path is as follows:

3.1.1. Teachers' subject teaching

It is possible to accurately diagnose the deficiencies in the teaching of the subject matter of the teacher through the teacher portrait model, preschool classroom teaching should pay attention to the design of diversified activities, to fully guide students to actively participate, to organize students to carry out group discussions, collaborative inquiry, and to organize activities oriented to complex problems. It is necessary to grasp the logic and law of teaching content, alternating between static and dynamic, and combining lecture and practice. To strengthen the inspirational teaching and thinking training, we should prepare more suitable resources for students and guide them to carry out independent inquiry learning, which not only improves students' theoretical knowledge, but also promotes the professional development of preschool teachers.

3.1.2. School Teaching and Research Organization

Organically combining teaching and research with the teacher portrait model, it focuses on helping teachers solve practical problems in specific teaching through intelligent diagnosis and lesson observation, realizing the integration of research and learning, and enabling teachers to learn, practice, apply and improve at the same time. Focusing on lesson examples, a number of high-quality teaching resources are selected, including teaching design plans, classroom teaching videos, teachers' lecture videos or lecture texts, classroom teaching resources, etc. Through the analysis of the intelligent teaching and research platform, teachers' discussions and evaluations, evaluation can be realized to promote research and study, as well as observing and learning, and seminars. Intelligent teaching and research must focus on the design of teaching activities to carry out teaching and research, and the design of teaching activities can reflect practicality, comprehensiveness, openness and inquiry as much as possible, and the teaching methods can reflect inspirational, inquiry, discussion and cooperative teaching as much as possible, aiming at improving teachers' professional level.

3.2. Validation of Preschool Teacher Development Pathways

In order to explore the effect of the application of preschool teacher development pathway that integrates teacher portrait model, so an empirical study was conducted on the course of third year undergraduate students of preschool education in a school. This study used empirical research, quantitative methods, and the quantitative research was mainly questionnaire rating scales. The following is a detailed description of the research design, research instrument and research model throughout the study.

3.2.1. Research hypothesis

The teacher portrait model centers on teacher characteristics and focuses on the process of teachers' professionalism improvement, mainly examining teachers' performance in teaching skills, theoretical knowledge of teaching and learning, as well as values. Therefore, the following research hypotheses are made before the study:

(1) The preschool teacher development path based on the teacher portrait model can improve teachers' teaching skills.

(2) The preschool teacher development path based on the teacher portrait model can improve teachers' theoretical knowledge of teaching.

(3) The preschool teacher development path based on the teacher portrait model can enhance teachers' values.

3.2.2. Research sample

In this study, 100 preschool teachers of a university were used as experimental subjects, and they were divided into experimental and control groups in parallel with the gender factor, in which the experimental group consisted of 50 persons, 19 male and 31 female. The control group consisted of 50 people, 25 male and 25 female. The experimental group was intervened with the preschool teacher development pathway based on the teacher portrait model, while the control group was intervened with the traditional preschool teacher development pathway, and other experimental cycles, environments, and other factors remained unchanged.

3.2.3. Research tools

To engage in empirical research of survey research, it is necessary to have the choice of experimental tools, and the correct experimental tools are the guarantee for the smooth progress of the experiment. By analyzing the pre- and post-test data of the experimental group and the control group, we can then verify the practical application value of the preschool teachers' development path in this paper. The quantitative tools are the following two:

(1) SPSS statistical software

Statistical methods are increasingly valued in research in the field of education. SPSS statistical software, as the world's earliest industry-leading statistical analysis software, has powerful advantages in data organization, data statistics and data analysis. With the help of this software, this study statistically analyzes the pre- and post-test data of the experimental group and the control group to explore the practical application effect of the preschool teacher development path in this paper.

(2) Scoring Criteria Scale

Combined with the existing research data, the professional level scale of preschool teachers is designed, which consists of three dimensions, namely, teaching skills, theoretical knowledge of teaching and values, and there are 10 items for each dimension, with scores from 1 to 5 indicating unsatisfactory, less satisfactory, generally satisfactory, satisfactory, and quite satisfactory, respectively. In order to make the results of the study more convincing, the scale was tested for reliability and validity, and it was found that the scale passed the reliability and validity test and could be used for subsequent research and analysis.

3.2.4. Findings

(1) Homogeneity test

Because all the subjects were teachers taken from a school with roughly the same basic information, the t-test of independent samples was conducted on their professional level of teachers before the implementation of the intervention, and the results of the homogeneity test are shown in Fig. 4, in which Q1~Q3 denote the teaching skills, theoretical knowledge of teaching and values respectively, and CG and EG denote the control and experimental groups respectively. As can be seen from the figure, there is no significant difference between the experimental group and the control group in terms of teaching skills, theoretical knowledge of teaching, and values. It can be seen that the two groups, the experimental group and the control group, are equal groups and can be directly subjected to the two independent samples t-test in the post-test and the follow-up test.

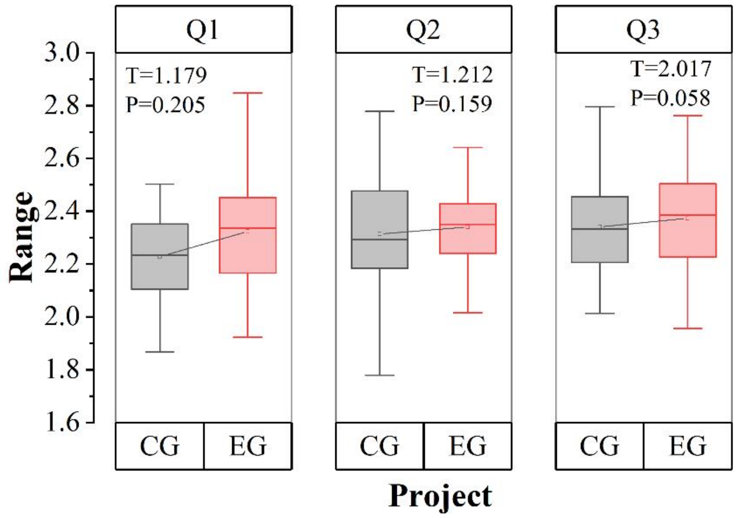

Figure 4. Homogeneity test results.

(2) Test of experimental and control groups on teachers' professional level after intervention

After the experimental intervention, the differences between the experimental group and the control group in various dimensions of teachers' professional level are compared to explore the application effect of the teacher portrait model based on data mining technology on the professional level of preschool teachers, and the results of the post-intervention comparative analysis are shown in Figure 5. It can be seen that in the post-test, there are significant differences between the experimental group and the control group in teaching skills, theoretical knowledge of teaching and values, indicating that the teacher portrait model based on data mining technology has a certain immediate counseling effect on teacher training.

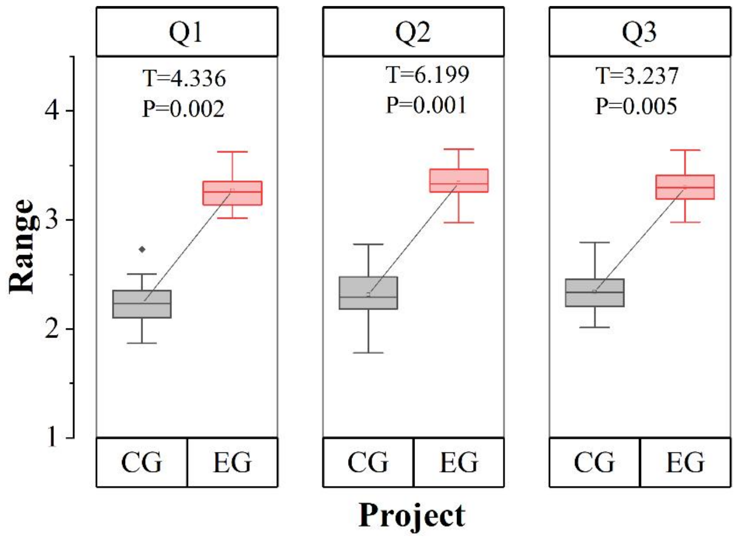

Figure 5. Comparative analysis after intervention.

(3) Comparison of the differences between the experimental group and the control group in communication skills on the tracking test

One month after the end of the intervention, the experimental group and the control group were tested on the professional level of teachers, and the differences between the experimental group and the control group on the tracking test were compared to explore the effect of continuous counseling of the Teacher's Portrait Model, and the results of the analysis of the differences are shown in Fig. 6. From the data performance in the figure, it can be seen that the differences between the experimental group and the control group in teaching skills, theoretical knowledge of teaching and values reached a significant level, which shows that there is a continuous counseling effect of the teacher portrait model in the improvement of teachers' professionalism.

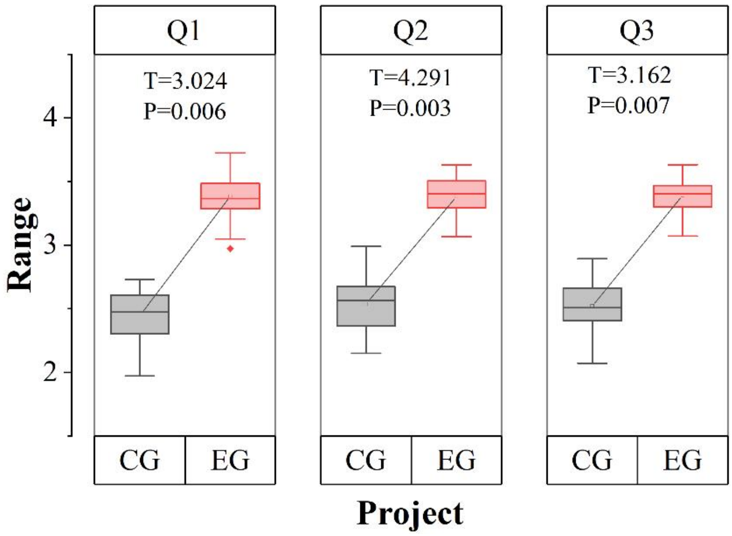

Figure 6. Results of the difference analysis.

(4) Comparison of the differences between the teacher's professional level in the experimental group on the pre-test and post-test

The data of the experimental group before the implementation of the experiment and after the implementation of the test were subjected to a one-sample t-test, in order to check, measure the effect of the experiment, the details of which are shown in the following Figure 7. From the figure below, it can be seen that the experimental group is in the teaching skills, teaching theoretical knowledge and values of the differences are significant level, the introduction of the teacher portrait model in the traditional preschool teachers' professional development path will help teachers to improve their professional level.

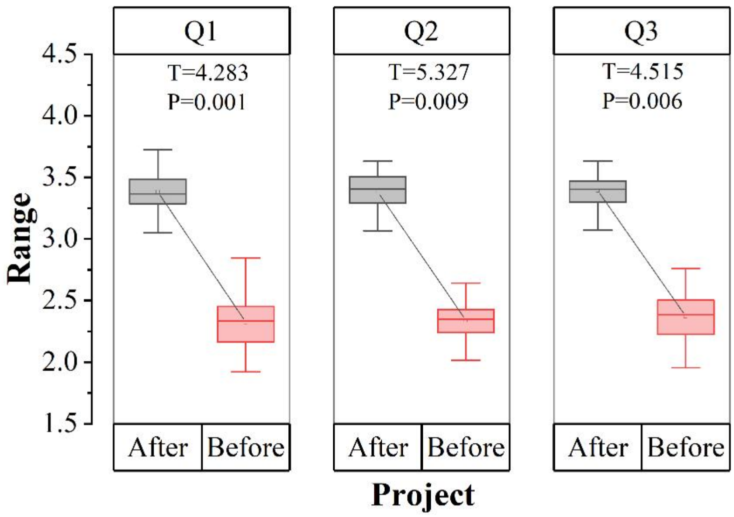

Figure 7. Comparison of differences.

4. Conclusion

The professional development of preschool teachers has been a long-standing concern in the education and academic sectors. This paper designs a teacher portrait model based on data mining technology under the perspective of intelligent education big data, and proposes a brand-new professional development path for preschool teachers with the help of the intervention of this model. Comprehensive scale testing and statistical analysis methods are used to validate and analyze the professional development path of preschool teachers integrating the teacher portrait model. The results of the study are as follows:

Under the effect of the model in this paper, the professional level of preschool teachers can be divided into “novice” preschool teachers, “skilled” preschool teachers, “expert” preschool teachers, “expert” preschool teachers, ‘expert’ preschool teachers and “expert” preschool teachers. The professional level of preschool teachers is divided into “novice” preschool teachers, “skilled” preschool teachers and ‘expert’ preschool teachers, in which “novice” preschool teachers (number: 235, proportion: 47%) are much larger than the other two types, which reflects the application of the model of this paper in the professional development of preschool teachers, and in addition to verifying the reliability of the model of this paper.

The differences between the experimental group and the control group in teaching skills ($T=4.336$, $p=0.002<0.05$), theoretical knowledge of teaching ($T=6.199$, $p=0.001<0.05$), and values ($T=3.237$, $p=0.005<0.05$) belong to the level of significance, i.e., compared with traditional paths, the preschool teacher professional development path that integrates the teacher portrait model is more conducive to the preschool teacher professional development. development path is more favorable to preschool teachers' professional development.

References

1. Qi, X., & Melhuish, E. C. (2017). Early childhood education and care in China: History, current trends and challenges. Early Years, 37(3), 268-284.
2. Pan, Y., Wang, X., & Li, L. (2018). Early childhood education and development in China. In International handbook of early childhood education (pp. 599-622). Dordrecht: Springer Netherlands.
3. Jiang, Y., Pang, L. J., & Sun, J. (2017). Early childhood teacher education in China. Early childhood education in Chinese societies, 85-100.
4. Li, K., Pan, Y., Hu, B., Burchinal, M., De Marco, A., Fan, X., & Qin, J. (2016). Early childhood education quality and child outcomes in China: Evidence from Zhejiang Province. Early Childhood Research Quarterly, 36, 427-438.
5. Liu, Y. (2023). Theoretical Exploration of Practice Teaching Reform in Preschool Education Using Social Construction Theory. people, 6(25), 142-146.
6. Song, Z., Zhu, J., Xia, Z., & Wu, X. (2014). The early childhood education of disadvantaged children in China. European Early Childhood Education Research Journal, 22(3), 355-365.
7. Li, H., & Liu, H. (2020, October). Preschool Education Model based on big data era. In 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI) (pp. 323-328). IEEE.
8. Liu, Y. Y. (2021). An Analysis of Preschool Teachers' Quality: Using Big Data as an Example. School Administrators, (134), 118-137.
9. Wu, R. (2022). Path of preschool education personnel training under the background of Big Data. International Journal of Educational Innovation and Science, 3(1), 1-9.
10. Meng, L., Xin, Q., & Fan, Q. (2022, July). Application of artificial intelligence in pre-school education professional talent training in the era of big data. In International Conference on E-Learning, E-Education, and Online Training (pp. 654-670). Cham: Springer Nature Switzerland.
11. Han, Y. (2021, September). Path Analysis of Pre-school Education Talents Training under the Background of Big Data. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 2554-2557).
12. Sun, Y. (2023, September). Research on Data Mining of Preschool Education Teachers’ Information Literacy Based on Association Rule Algorithm. In 2023 International Conference on Telecommunications, Electronics and Informatics (ICTEI) (pp. 131-134). IEEE.
13. Gafarov, F. M., & Galimyanov, A. F. (2020). Identification of Key Factors in the Formation of an Individual Trajectory of Teacher Professional Developmentin Digital Environment Based on Big Data. ARPHA Proceedings, 3, 625-632.
14. Leskina, I. (2022). Model for managing the professional performance of a teacher through the use of big data. Man and Education, (4 (69)), 105-116.
15. Besbes, R., & Besbes, S. (2016, March). Cognitive dashboard for teachers professional development. In Qatar Foundation Annual Research Conference Proceedings (Vol. 2016, No. 1, p. ICTPP2984). Qatar: HBKU Press.
16. Cui, H., & Zhang, D. (2018, April). Strategies on teacher professional development in big data era. In 2018 4th International Conference on Education Technology, Management and Humanities Science (ETMHS 2018) (pp. 324-328). Atlantis Press.

17. Sun, H., Zhang, H., & Zheng, Q. (2018, December). Data Model Building for Online Teachers' Professional Development. In 2018 Seventh International Conference of Educational Innovation through Technology (EITT) (pp. 250-253). IEEE.
18. Jiapin Cao. (2025). Analysis and New Connotation of the Improvement Path of College Teachers' Quality using Data Mining in the Perspective of Establishing Virtuous People. Applied Mathematics and Nonlinear Sciences,10(1),
19. Min Yang. (2024). Construction of Teacher Personal Knowledge Management Model Based on Data Mining Algorithm from TPACK Perspective. Information and Knowledge Management,5(1).
20. Hu Xingcui. (2023). The evaluation method of college teachers' morality considering intelligent emotion recognition and data mining algorithm. Applied Artificial Intelligence,37(1).
21. Ordoñez-Avila Ricardo,Salgado Reyes Nelson,Meza Jaime & Ventura Sebastián. (2023). Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review. Heliyon,9(3),e13939-e13939.