ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

Teaching Research of the Combination of Educational Statistics and Computer Based on Big Data Fusion

Wei Liu

School of Statistics, Shandong Technology and Business University, Yantai, Shandong, 264100, China
sunshineliuwei0168@163.com

Abstract: *This study aims to explore the teaching method of combining educational statistics with computers based on big data fusion, aiming to improve the quality of education and optimize the allocation of teaching resources. The article first analyzes the profound impact of computer technology on the field of education in the era of big data, and points out that the teaching design combining educational statistics with computers has become an important research direction at present. Then, the characteristics of big data and its transformative role in the education industry are elaborated in detail, and it is proposed to design an intelligent teaching platform through big data technology. The study adopts a decision tree algorithm to ensure the consistency of system selection and user information based on the matching of user needs and teaching resources. In addition, the study also uses unsupervised learning methods to perform cluster analysis on user information, and realizes self-optimization of user information through the semantic description of ontology. Finally, the effectiveness of the method is verified through a series of teaching experiments, and the results show that the platform can provide a more personalized learning experience and teaching resource configuration. This study provides new ideas for educational practice, especially in the field of teaching combining educational statistics with computer technology, which has important reference value.*

Keywords: *Big Data, Educational Statistics, Computer Teaching, Teaching Research*

1. Introduction

The 21st century is an important era with the rapid development of computer and network technology. Facing the opportunities and challenges brought by the comprehensive informatization of society, the education industry cannot lag behind in this regard. Education has always been a topic highly valued by the state and society. With the development of human civilization for thousands of years, education has been improving with the development of the times. As one of the important cognitive tools for the development of the new era, computer technology will develop into an important curriculum auxiliary tool in this century. The combination of computer technology and school classroom is expected both at home and abroad.

The rapid development and wide application of geographic information system (GIS) and the advent of big data era have brought the emergence of multi-source spatial data. Due to the differences of data sources, data accuracy and data modes, data integration and fusion sharing are more difficult. Huang Z proposed the formation and development trend of multi-source spatial data fusion methods. Based on consulting a large number of relevant technical documents at home and abroad, he summarized and discussed the multi-source spatial data fusion methods, which has a certain reference value for relevant research work [1]. Online teaching and face-to-face teaching are different experiences. The knowledge and skills developed for face-to-face teaching are insufficient to prepare for online teaching. Teaching science, technology, engineering and Mathematics (stem) courses entirely online is more challenging because these courses usually require more practical activities and live demonstrations [2]. Digital story telling (DST) has recently emerged as a new teaching tool in the teaching environment. DST combines media and technology with traditional storytelling to help students learn. Kocaman Karoglu an examined the use of DST in university courses and the views of pre service teachers on their learning experience of using the tool. 38 pre service teachers participated in the study, in which DST was used as a learning activity. In the course, participants created personal digital stories about their careers [3]. Scholars have never been slack in the research of teaching methods. The purpose of the Kmen B study is to define the language teaching methods used by Turkish

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

English teachers and their use level, and to determine whether the use level varies according to gender, qualification and type of graduation school. The research group consists of 95 English teachers who studied in Duzce middle school from 2013 to 2014 [4]. Statistics has always been a subject that researchers cannot ignore. Ruz f studied people's attitudes towards descriptive statistics and its teaching with 126 quasi mathematics teachers from Spain and Chile as samples. Following the quantitative approach, we analyzed in depth the attitudes (negative or positive) declared by participants. Firstly, he refers to the validity of the scale, then discusses the configuration of the sample to different attitude components, and finally ends according to the main differences in the countries / regions where the participants are located [5]. The field of educational technology has always been one of the hot spots of social concern. Ritzhaupt A D records the design, development and evaluation of statistical courses tailored for Educational Technology Doctoral Students in online courses, which aims to help online doctoral students use relevant and real learning experience to conduct consumption and quantitative research in the field of educational technology. He first outlined the curriculum features used to engage students in learning materials [6].

This paper establishes a network teaching platform mainly dominated by students, analyzes and models students' information during learning activities, selects the way of decision tree to judge the consistency between users and user information, provides more teaching resources, and uses unsupervised learning for information clustering on the basis of user information labels studied by predecessors, Through the semantic description based on ontology, the self-improvement of user information is realized.

2. Teaching Research Method of Combining Educational Statistics and Computer Based on Big Data Fusion

2.1 Characteristics of Big Data

With the rise of major network platforms, the amount of data in the world has increased significantly. Table 1 shows the changes in the amount of data from 2014 to 2018 [7].

Table 1: Changes in data volume from 2014 to 2018

particular year	2014	2015	2016	2017	2018
Changes in data volume	460	700	1026	1990	3290
Increased data volume	-	240	326	964	1300

Big data has had an important impact on the education industry. When researchers found that the substantial growth of big data has deeply affected the whole society, it has become an urgent problem to enable the whole society to quickly adapt to this new era. In 2018, more than 200 colleges and universities have approved the application for adding majors related to big data technology. Figure 1 shows the distribution statistics of colleges and universities majoring in big data in 2018, in which arts, science and engineering increased the most [8].

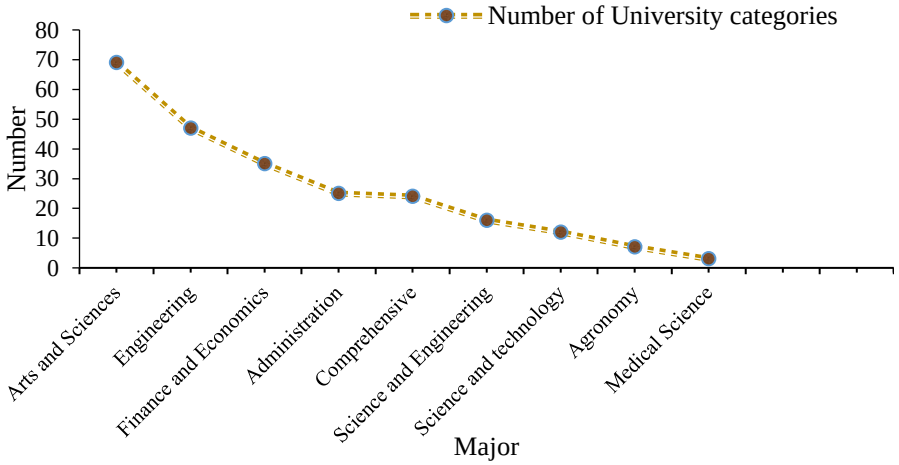

Figure 1: Distribution statistics of universities majoring in big data in 2018

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

2.2 Design of Teaching Platform Based on Big Data Technology

Students are one of the important roles in learning activities. It is particularly important to provide each student with a fully functional and resource rich platform. In order to realize the personalized and individualized teaching system structure, we must effectively understand the needs of students. In this regard, we need to compare and select according to the experience and data of many previous research scholars, and establish a strong teaching resource database. In the teaching method of combining educational statistics with computers, teachers mainly conduct guided learning. Teachers can guide students according to the platform records, or appropriately select students to study independently by using the teaching platform. Teachers' participation seems to be less, but it is important to always pay attention to the learning status and progress of each student [9]. The teaching platform system can push teaching resources with different difficulties according to each student's cognitive ability and learning ability, which reduces the embarrassing situation that teachers need to face. Some students have mastered it, and some students still look ignorant[10].

Information consciousness is people's dynamic response to information, which is manifested in their keen perception of information, correct observation, analysis and judgment of information value and innovation of information. The cultivation of information awareness starts from two aspects. We need to guide students out of the classroom, collect information through statistical investigation, eliminate the false and preserve the true, obtain valuable information through sorting and summary, and improve students' sensitivity to information data [11].

(1) The process of retrieving students' needs

Since the demand of platform users for teaching resources is dynamic, we need to be sure to perceive accurately according to the query conditions, but we also need to infer and select according to the changing user needs [12].

When the learning user needs to use the system platform, the user account starts to log in. Platform users start learning according to their own needs. The platform starts to retrieve user needs, extract previous learning records for auxiliary query, match the teaching resources most suitable for users, and collect all user information at the current time t as. At the same time, users can evaluate the pushed teaching resources in time, so as to update the teaching resources as soon as possible [13].

(2) Understanding and learning user information

Without understanding and learning user information, all subsequent query activities are futile. The system needs to match the required teaching resources according to the user information. The following describes in detail the two methods of obtaining and identifying user information by directly obtaining user information and associated obtaining user information [14].

Direct access to the user's information is integrated through the decentralized information input by the user, and then stored.

Association acquisition of user information is the direct acquisition of information elements, system active reasoning and association of edge resources and constraint set. In determining the composition of the user model, the most appropriate correlation function and results are determined according to the system's automatic correlation and reasoning, so as to collect user information and improve the user's learning experience [15].

(3) Decision tree algorithm

In this paper, the decision tree is used to determine whether the system selection is consistent with the user information. In the computer field, the decision tree is composed of three important nodes: primary root node, intermediate node and leaf node [16]. The user known information system classifies the integrated user known information according to different attributes. At the position of the primary root node, the result attribute of each analysis object has a node corresponding to it. At this time, the primary node is further divided into a subtree, including a new decision path and the path tree corresponding to the attribute combination of a node. Starting from the primary node, it finally reaches the child node through some key paths and nodes. There is only one path to ensure that the output of the decision tree is unique [17].

Therefore, this paper uses the decision tree method to determine the consistency with user information. In order to detect whether the teaching resources given by the platform match, we need to compare the known user information with the user information base in the database. The known user

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

information is recorded, and the user information in the database is recorded. The user information similarity is obtained by constructing the decision tree for node operation, and the user information model tree is constructed [18].

Therefore, to calculate the user information similarity, first find the nodes of the tree graph corresponding to I_1 and I_2, and then calculate the similarity of the corresponding nodes.

1) Find the corresponding node. If V is a node in the user information model tree, find (V, V ') [19] the algorithm for finding the corresponding node in the user information model tree in the database.

For node similarity calculation, if the learning achievement node is recorded as S, the attribute set it contains is recorded as T, $T = \{t_1, t_2, t_3, ..., t_n\}$, and the corresponding node of its user information model tree I_2 is $I^{'}$, the similarity calculation formula of the two nodes is:

$$sim(S, I^{'}) = \sum_{i=1}^{n} w_j \cdot simP_i(S, I^{'}) \tag{1}$$

$simP_i(S, I^{'})$ in expression (1) Indicates the similarity of the i-th attribute between node S and the corresponding node $I^{'}$, and u_i indicates the weight value of the attribute [20]. The character attribute is expressed as CHP, the numerical attribute is expressed as NP, the object attribute is expressed as OP, the Category attribute is expressed as CP, and the similarity calculation formula of single value character attribute is [21]:

$$simT_{(CHP)}(chp_1, chp_2) = \begin{cases} 1(chp_1 \cdot value = chp_2 \cdot value) \\ 0(\text{Other}) \end{cases} \tag{2}$$

The calculation formula of single value numerical attribute similarity is:

$$simT(NP)(np_1, np_2 = \frac{\| np_1 | - | np_2 \|}{| np_1 |}) \tag{3}$$

The calculation of attribute similarity of single valued object is equivalent to the nested calculation between user information elements [22], so the calculation formula is:

$$sim_{(OP)}((op_1, op_2) = \sum_{i=1}^{n} u_j \cdot simT_i(op_1, op_2) \tag{4}$$

When calculating whether the single value category matches, the single value classification hierarchy attribute takes a tree with edge weight composed of node set A and edge set E, l represents the length of the longest path in the tree, and T (b1, b2) represents the distance between nodes. Therefore, the similarity calculation formula is [23]:

$$sim_{cp}(cp_1, cp_2) = \frac{L - | \mathrm{T}(cp_1, cp_2) |}{L} \tag{5}$$

When calculating the consistency of multivalued attributes, each attribute value set is expressed as g. to calculate the consistency of two sets, first calculate the consistency of elements and sets, and the formula is:

$$sim(s, G) = \max(sim_s(s, s_i) \mid s_i \in G) \tag{6}$$

Use | G | to represent the number of elements in the set, Calculate the consistency of two sets. First, calculate the consistency of elements and sets. The formula is:

$$sim(G_1, G_2) = \frac{(\sum_{s \in G_1} sim(s, G_2) + \sum_{s \in G_2} sim(s, G_1))}{| G_1 | + | G_2 |} \tag{7}$$

When we determine the calculation method of relevant attributes, we also need to determine the

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

weight distribution to be closer to the user needs. The weight distribution needs to change the user information and update the user needs at any time [24]. Therefore, the user's needs are expressed from three dimensions: query criteria, filter rules and sorting rules. The user's teaching resource needs are expressed as k = {QC, EF, or}. QC represents the set of teaching resource query criteria, EF represents the set of filter rules, and or represents the sorting rules of teaching resources. The query criteria matrix is:

$$\begin{bmatrix} P_1(p_{11}, p_{12}, \ldots\ldots, p_{1n}) \\ P_2(p_{21}, p_{22}, \ldots\ldots, p_{2n}) \\ \ldots\ldots \\ P_m(p_{m1}, p_{m2}, \ldots\ldots, p_{mn}) \end{bmatrix}, m > 0, n > 0 \tag{8}$$

When the learning user starts learning activities, querying teaching resources is the system to collect the teaching resources required by the user with the resources corresponding to the query criteria QC. Among them, the corresponding teaching resources after each restriction condition entered by the user also need to be collected. Therefore, if B (P) represents the teaching resource set corresponding to the query condition, and Q (P) represents the teaching resource set corresponding to each query condition, the teaching resources that the platform user needs to query can be expressed as:

$$Q(QC) = U_{i=1}^{m}(Q(Q_i)) = U_{i=1}^{m}(\cap_{j=1}^{n} q(q_i)), m > 0, n > 0 \tag{9}$$

1) EF refers to the rule set for filtering the query criteria, and then the queried teaching resources need to be filtered to delete unnecessary teaching resources. Each deleted teaching resource set is recorded as Q (EF), and the final deleted set is recorded as B (EF). Then the deleted large data set can be expressed as

$$q(EF) = U_{i=1}^{m}(q(ef_1)), m > 0 \tag{10}$$

2) Or represents the sorting rule of the results of teaching resource demand, weights the teaching resource result set formed after query and screening through the preset weight distribution rules, and sorts the teaching resource demand according to the weight results. The weighted rule set is expressed as the teaching resource result set formed after query and screening, and is expressed as a. Finally, the big data demand result is expressed as B [25]

$$L(q) = (\coprod_{i=1}^{m} L_i(q_1), \coprod_{i=1}^{m} L_i(q_2), \ldots, \coprod_{i=1}^{m} L_i(q_n)), m > 0, n > 0 \tag{11}$$

After the hierarchical setting of query conditions, deletion rules and weight rules for user requirements, the representation method of user teaching resources has been determined. The query conditions QC of teaching resource requirements are constructed according to the learning process, learning objects and learning tasks in user information elements. The construction process is as follows:

$$\begin{Bmatrix} P \\ O \\ T \end{Bmatrix} \overset{yields}{\Rightarrow} QC \tag{12}$$

3) The teaching resource demand query conditions are constructed according to the learning object, and finally the teaching resource demand set qCO according to the learning object is obtained [26].

3. Teaching Method Experiment of Combining Educational Statistics and Computer Based on Big Data Fusion

This survey first selected two schools as the survey object. In the process of sample selection, the method of random stratified sampling was adopted to make the sample representative as far as possible. A total of 300 questionnaires were distributed and 270 were recovered. After excluding the invalid questionnaires, the number of valid questionnaires was 270. The content of the questionnaire set up

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

different views on the teaching research of the combination of educational statistics and computer.

In this paper, two classes with almost equal basic quality in the same school are selected for the experiment for one month. Class a adopts the traditional teaching method and class b adopts the teaching method combined with computer. One month later, the differences of teaching methods are compared by comparing the learning achievements and learning status of the two classes.

4. Experimental Results and Analysis of Combined Teaching of Educational Statistics and Computer Based on Big Data Fusion

Figure 2 shows the survey number of students in five majors to study educational statistics. The majors surveyed include pedagogy, psychology, educational management, primary education and preschool education. Among them, pedagogy and educational management are the most investigated majors. Figure 3 shows the number of freshmen, sophomores and juniors in the survey.

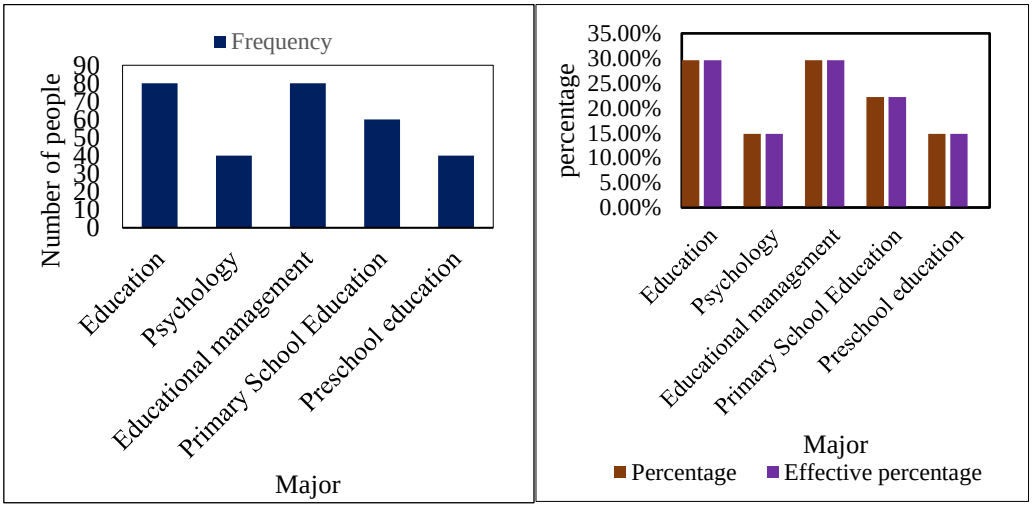

Figure 2: Survey number of students studying five majors of Educational Statistics

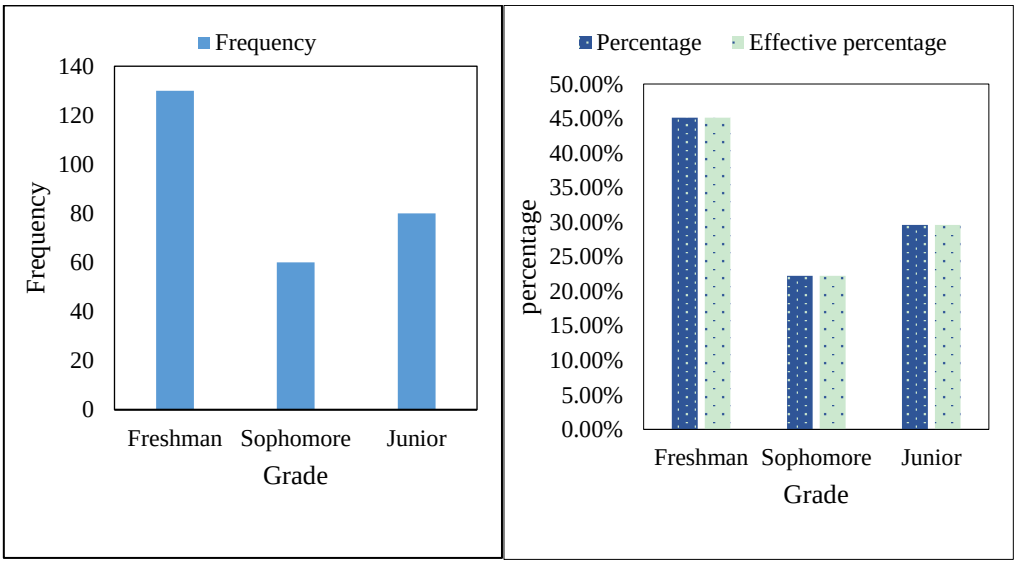

Figure 3: The number of freshmen, sophomores and juniors in the survey

Table 2 shows the number of male and female students in the survey. Figure 4 shows the students' views on the teaching method of combining educational statistics with computer. 44.4% are in favor and 88.8% are not against it. Only when students do not reject this teaching method, can teaching research be carried out smoothly.

Table 2: Number of male and female students in the survey

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

	Male	Female
Frequency	140	130
Percentage	54.86%	45.14%
Effective percentage	54.86%	45.14%

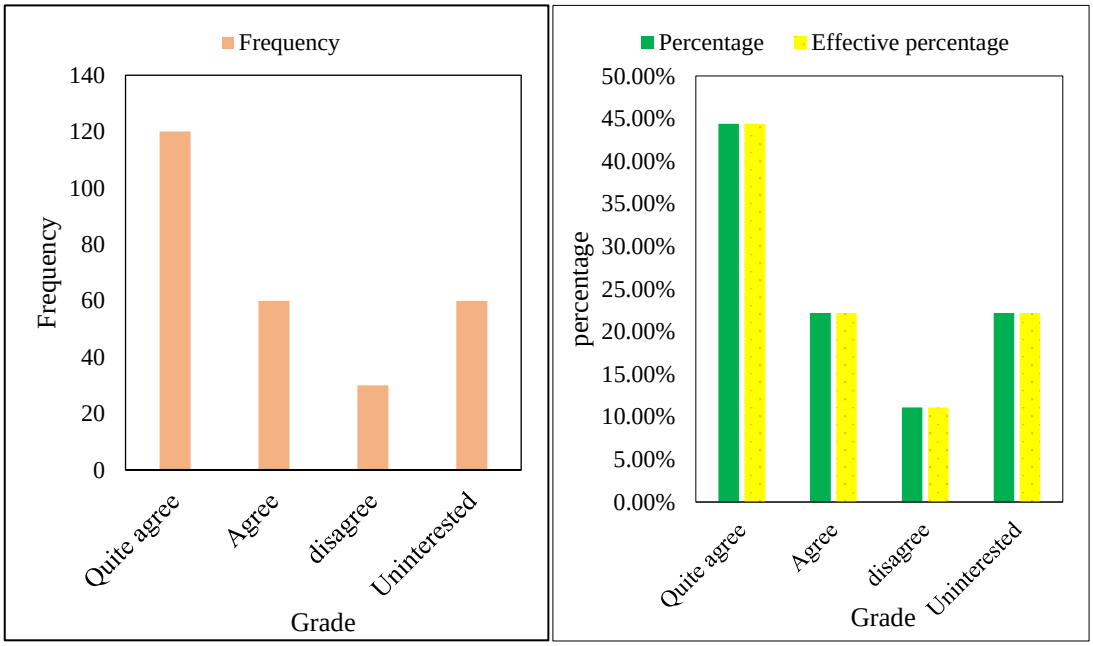

Figure 4: Students' views on the teaching method of combining educational statistics with computers

Table 3 and figure 5 show the academic performance and learning status of two experimental classes a and b. On the whole, class b students perform better in terms of academic performance (as can be seen from the pass rate and excellent rate) and classroom learning habits (whether to review before and after class, actively communicate with classmates and teachers, submit homework, attend classes on time, etc.). It is worth noting that the number of students in class b taking notes on their own initiative is not optimistic, which may be far from the complete functions of the online platform or the distance between teachers and students, resulting in some slack in students' thinking.

Table 3: Number of pre class preview and post class review

	a	b
Preview before class	30	40
Review after class	40	45
Hands up in class	30	45

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

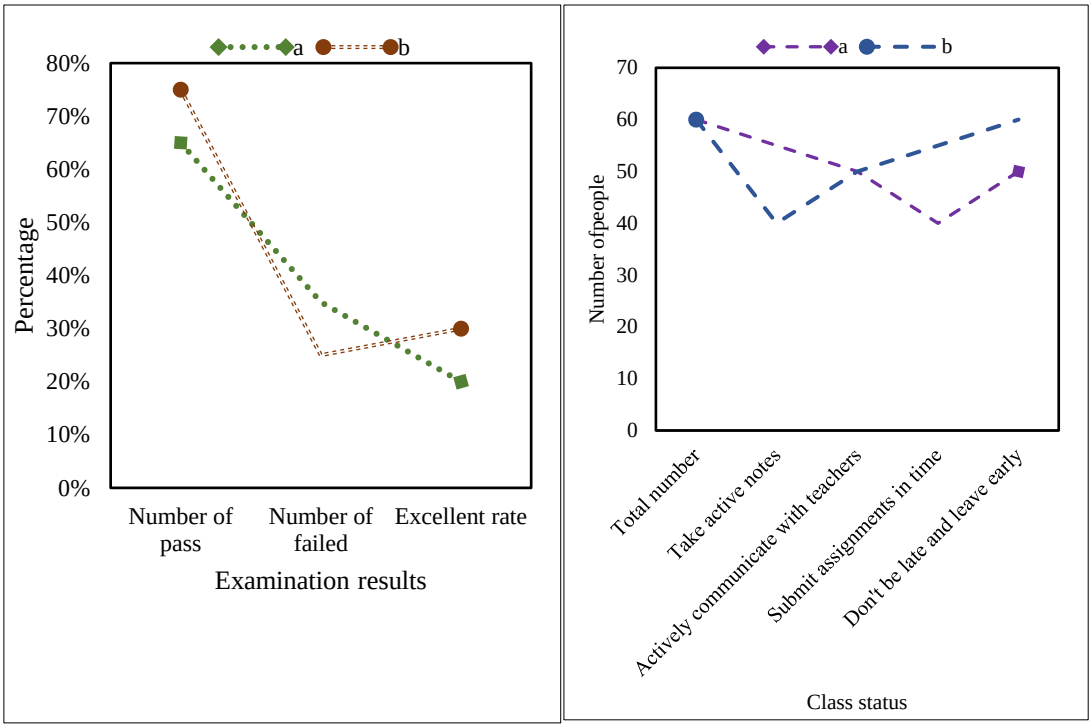

Figure 5: Academic performance and learning status of two experimental classes a and b

Table 4 shows the views of class b students on the role of the use of the online platform in promoting the learning effect, and table 5 shows the students' effective use of the resources of the online platform. Among them, 90% of the students think that the use of the network platform can promote the learning effect. In terms of the resource utilization of the network platform, they actively participate in the aspects that students are interested in, such as watching teaching courseware, answering common questions, doing simulated test questions and testing.

Table 4: Views of class b students on the role of the use of network platform in promoting learning effect

	Number of people	percentage
Very good function	30	50%
Better effect	15	25%
Generally good	9	15%
No effect	5	8.3%
Negative effects	1	1.7%

Table 5: Class b students' effective use of online platform resources

	Course case analysis	Course forum	Frequently asked questions	Simulated test questions	Online test	Teaching video	Teaching courseware
Number of people	5	20	55	45	43	15	60
Percentage	8.3%	66.7%	91.7%	75%	71.7%	25%	100%

5. Conclusion

This paper studies the teaching of educational statistics and computer based on big data fusion, which is of great practical significance to the education industry. At present, the research result is that the combination of educational statistics and computer has great advantages, but there are still many technical problems to be broken through in the specific implementation. At the same time, there are still many deficiencies in some aspects adopted in this paper. As follows: (1) refinement of user information. Due to the strong sparsity of big data, when users learn on the platform, the lack and inconsistency of data in the process of data collection lead to a high degree of lack of user information. Therefore, the

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

collection and analysis of user information need to be further improved. (2) Improvement of decision algorithm. The decision tree has no backtracking mechanism when the node splits and selects an attribute for decision analysis, so it is easy to achieve local optimization in user model matching, but it can not achieve global optimization. Therefore, it is necessary to further study and optimize the decision algorithm.

References

[1] Huang Z, Fu Y, Dai F. Study for Multi-Resources Spatial Data Fusion Methods in Big Data Environment [J]. Intelligent automation and soft computing, 2018, 24(1):29-34.
[2] Dazhi, Yang. Educational research Instructional strategies Online course design Online pedagogy Statistics Teaching STEM online[J]. International journal of STEM education, 2017, 4(1):34-34.
[3] Kocaman-Karoglu A. Personal voices in higher education: A digital storytelling experience for pre-service teachers[J]. Education & Information Technologies, 2016, 21(5):1-16.
[4] Kmen B, Kl A. A Research about the Level of Using Language Teaching Methods and Its Effect on Some Variables: in Turkey[J]. Universal Journal of Educational Research, 2016, 4(9):1994-2001.
[5] Ruz F, Molina-Portillo E, Contreras J M. Attitudes Towards Descriptive Statistics And Its Teaching In Prospective Teachers[J]. Cadernos de Pesquisa, 2020, 50(178):964-980.
[6] Ritzhaupt A D, Valle N, Sommer M. Design, Development, and Evaluation of an Online Statistics Course for Educational Technology Doctoral Students: a Design and Development Case[J]. Journal of Formative Design in Learning, 2020, 4(2):119-135.
[7] Liu X, Wang W, Zhu G. Research and analysis of big data based on hadoop[J]. Boletin Tecnico, 2017, 55(4):382-386.
[8] Tnsing K M. Supporting the Production of Graphic Symbol Combinations by Children with Limited Speech: A Comparison of Two AAC systems[J]. Journal of Developmental and Physical Disabilities, 2016, 28(1):5-29.
[9] Cohen-Mimran R, Reznik-Nevet L, Korona-Gaon S. An Activity-Based Language Intervention Program for Kindergarten Children: A Retrospective Evaluation[J]. Early Childhood Education Journal, 2016, 44(1):1-10.
[10] Ben-Zvi D, Makar K. A Framework for Assessing Statistical Knowledge for Teaching Based on the Identification of Conceptions of Variability Held by Teachers[J]. The Teaching and Learning of Statistics, 2016, 10.1007/978-3-319-23470-0(Chapter 37):315-325.
[11] Umugiraneza O, Bansilal S, North D. Investigating teachers' formulations of learning objectives and introductory approaches in teaching mathematics and statistics[J]. International Journal of Mathematical Education in Science & Technology, 2018, 49(7-8):1148-1164.
[12] Saglimbene V, Strippoli G, Craig J C, et al. Statistics and data analyses—a new educational series for nephrologists - ScienceDirect[J]. Kidney International, 2020, 97(2):233-235.
[13] Ho, A. D. The New (Educational) Statistics: Properties of Scales That Matter[J]. Journal of Educational & Behavioral Statistics, 2016, 41(1):94-99.
[14] Carl M. Excel 2013 for educational and psychological statistics: a guide to solving practical problems [J]. Computing Reviews, 2016, 57(10):595-596.
[15] Nye J, Bryukhanov M, Polyachenko S. Descriptive statistics and regressions of 2D:4D and educational attainment based on RLMS data[J]. Data in Brief, 2017, 12(C):552-583.
[16] George, B, Macready, et al. The Use of Probabilistic Models in the Assessment of Mastery[J]. Journal of Educational Statistics, 2016, 2(2):99-120.
[17] Dennis, M, Roberts, et al. Reliability and Validity of a Statistics Attitude Survey[J]. Educational and Psychological Measurement, 2016, 40(1):235-238.
[18] Mcneish D M. Using Data-Dependent Priors to Mitigate Small Sample Bias in Latent Growth Models: A Discussion and Illustration Using Mplus[J]. Journal of Educational & Behavioral Statistics, 2016, 41(1):7-8.
[19] Bolsinova M, Tijmstra J. Posterior predictive checks for conditional independence between response time and accuracy[J]. Journal of Educational & Behavioral Statistics, 2016, 41(2):123-145.
[20] Tutz G, Berger M. Response Styles in Rating Scales: Simultaneous Modeling of Content-Related Effects and the Tendency to Middle or Extreme Categories[J]. Journal of Educational & Behavioral Statistics, 2016, 41(3):239-268.
[21] Liu Y, Tian W, Xin T. An Application of M2 Statistic to Evaluate the Fit of Cognitive Diagnostic Models[J]. Journal of Educational and Behavioral Statistics, 2016, 41(1):3-26.
[22] Mistler S A, Enders C K. A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data[J]. Journal of Educational and Behavioral Statistics, 2017, 42(4):432-466.

ISSN 2663-8169 Vol. 7, Issue 1: 71-80, DOI: 10.25236/IJNDE.2025.070111

[23] Ackerman, T. Discussion of David Thissens Bad Questions: An Essay Involving Item Response Theory[J]. Journal of Educational & Behavioral Statistics, 2016, 41(1):90-93.
[24] Naumann A, Hartig J, Hochweber J. Absolute and Relative Measures of Instructional Sensitivity[J]. Journal of Educational and Behavioral Statistics, 2017, 42(6):678–705.
[25] Jeon M, Boeck P D, Linden W. Modeling Answer Change Behavior: An Application of a Generalized Item Response Tree Model. [J]. Journal of Educational and Behavioral Statistics, 2017, 42(4):467-490.
[26] Zhang J, Wang J S, Du W, et al. Big data analysis reveals the truth of lumbar fusion: gender differences[J]. Spine Journal, 2017, 17(5):754-755.