

Research on the Factors Influencing the Academic Output of Outstanding Teachers Based on XGBoost

Meihang Guo^{1*}, Linan Lin¹

¹College of Preschool and Primary Education, Shenyang Normal University, Shenyang, China

*Corresponding Author. Email: 1982326219@qq.com

Abstract. This study uses the xgboost regression model to explore the main factors affecting the academic output level of good teachers. Based on the research, this paper selects the data from 123 outstanding teachers in the Beijing-Tianjin-Hebei area, with the selected variables including administrative experience factors, expert training factors and awards received factors. The number of variables affecting the academic output is seven. The results show that 50.20% of academic output is most significantly influenced by expert training experience, while 30.80% and 17.50% are affected by the number of awards and administrative experience, respectively. From the study results, it can be known that the positive influence of training experience, academic honors, and administrative roles on teachers' academic output has provided educational management departments with scientific basis for correct assessment and improvement of teacher resource allocation.

Keywords: XGBoost, Outstanding Teachers, Academic Output

1. Introduction

The Outline of the Plan for Building a Strong Educational Nation (2024–2035) guides governments at all levels to focus on the development of the teaching workforce, improve the teacher education system, enhance the quality of normal education, strengthen international exchange and cooperation in teacher training, and optimize teacher management and resource allocation, with the goal of building a high-quality and professional teaching staff. From the perspective of “promoting the spirit of great educators,” changes in the external environment have raised the issue of how to accurately identify and cultivate outstanding teachers. In particular, the exploration of the mechanisms affecting research output remains an important topic in improving teachers’ professionalism.

Current research on the cultivation of excellent teachers mainly involves cultivating high-quality teachers in terms of teacher professional development and improvement of teaching capabilities, as well as cultivation through teacher evaluation mechanisms. scholars in general hold the view that via different types of training like lectures and mentor-mentee coupling young teachers can be efficiently guided to rapidly adapt to the environment and work of education and teaching, improving their own professional standards. Some started to look at the output of good teachers academically, pointing out that traits like teachers' academic background, what they're researching interest, the things they do for a job, and the ability to innovate can directly change how much and how great that output is. As for the external environment in school such as research atmosphere in school, resources, cooperation resource, and policy and so on, also have significant impact on teacher’s academic output. For example, a strong academic atmosphere can stimulate teachers’ enthusiasm for research, abundant

resource support and broad cooperation opportunities can provide strong guarantees for academic research, while reasonable policy incentives can further promote the production and transformation of academic achievements [1]. But even though there have already been some research results on the academic output of excellent teachers, most of them focus on analyzing the factors that determine this output, and relatively few studies specifically deal with the scientific evaluation of academic output.

Academic output, as an important indicator of teachers' academic level and influence, not only reflects their academic ability and innovative spirit but also, to some extent, reflects the quality of education and research atmosphere in schools. With the development of artificial intelligence, machine learning has good application prospects. Compared to traditional statistical models and shallow machine learning methods, XGBoost, with its gradient boosting mechanism, regularization strategies, and parallel computing capabilities, can efficiently process multidimensional feature data of teachers (such as research input, teaching practices, team collaboration, policy support, etc.), uncover potential nonlinear relationships, and rank feature importance, thus overcoming the limitations of "experience-driven" evaluation. By using the XGBoost algorithm, schools can more accurately identify teachers with academic potential, optimize teacher resource allocation, improve overall academic standards, and scientifically formulate strategies for talent recruitment and development, thus enhancing the academic competitiveness and social influence of universities. For individual teachers, understanding the factors influencing their academic output allows them to make more targeted career development plans and improve academic output efficiency. This provides data-driven decision support for educational management departments, helping transform teacher training from "extensive management" to "precise intervention," ultimately serving the policy goal of building a high-quality teaching workforce.

2. Literature review

Scientific output, in simple terms, refers to the results achieved by researchers through scientific research activities within a certain period. These outcomes can take various forms, such as published papers, patent applications, published books, research project initiation and completion, and awards for research achievements. In the field of education, teachers' scientific output not only reflects their personal academic abilities and professional qualities but is also one of the important indicators for assessing their teaching level and contributions. For outstanding teachers, their research output is often more abundant and of higher quality, providing valuable academic resources and intellectual guidance for students and the education sector [2].

In the existing literature, scholars have different views on the factors influencing teachers' research output, focusing mainly on aspects such as school resources and environment, subject characteristics, personal social capital, educational philosophy and professional qualities, as well as research evaluation and educational systems. Hu Jingyun pointed out that factors such as school differences (e.g., "Double First-Class" universities versus regular universities), subject differences (e.g., science, engineering, and management), the relationship between research achievements and supervisors, the frequency of attending academic conferences, and the frequency of reading books or literature all significantly impact graduate students' research output. These factors also apply to teachers, especially school resources, academic atmosphere, subject characteristics, and support from research teams, which significantly affect teachers' research output. Zhou Xia believes that social capital factors such as overseas study experience, administrative positions, academic part-time jobs, and titles have a significant impact on both the quantity and quality of research output. Social capital factors provide teachers with more academic resources and opportunities to improve the quality of academic activities, which enhances teachers' academic influence and the ability to collaborate, and promotes the improvement of research output. Li Yu's research mainly focuses on the research and teaching

abilities of primary school teachers, and it also covers factors affecting the research abilities of primary school teachers. Teachers' educational philosophy, professional development awareness, and teaching experience are also factors that affect a teacher's research and teaching ability, and they indirectly affect research output through the support of schools for teaching and research. The above are also valid for secondary school and university teachers, indicating that teachers' personal attributes as well as the school environment have an impact on the quantity of their research output. Yang Mengting, analyzed the factors affecting the number of original research achievements in China, from historical and current aspects. She pointed out that cultural traditions, scientific reserves, research evaluations and educational systems are related to the production of original research results.

In view of the above background, this study selects seven factors—administrative experience, expert training, number of awards, collaborative ability, overseas study and exchange experience, excellent teaching examples from the Teaching and Research Network, and practical experience—to construct a model and propose an accurate evaluation model based on the XGBoost (Extreme Gradient Boosting) algorithm for scientific evaluation of outstanding teachers' academic output. XGBoost, as an efficient gradient boosting algorithm, performs excellently in classification, regression, and other tasks, with its performance and stability widely recognized. In the field of education, XGBoost can handle large-scale educational data, uncover deep-level associations, and is capable of processing high-dimensional data, nonlinear relationships, and missing values. Therefore, it is particularly suitable for teacher evaluation in complex educational scenarios. This model provides scientific support for teacher training and evaluation and helps address some issues in the existing evaluation system. The research can further optimize the XGBoost model by integrating more dimensions of data, improving the model's interpretability and stability, and providing more comprehensive and accurate support for educational decision-making. This not only helps promote teachers' personal growth and professional development but also contributes to the overall improvement of educational quality, creating more favorable conditions for students' all-round development.

3. Research design

This study aims to explore the factors influencing the academic output of 123 outstanding teachers in the Beijing-Tianjin-Hebei region, using XGBoost as a regression prediction model. The data for this study is sourced from the “Decision on Commending National Outstanding Teachers and National Outstanding Education Workers” issued by the Ministry of Education. The study selects features such as administrative experience, expert training, number of awards, collaborative ability, overseas study and exchange experience, excellent teaching examples from the Teaching and Research Network, and practical experience, with scientific output as the outcome variable. The research process primarily includes three main stages: data preprocessing, model construction and training, and model evaluation.

3.1. Data preprocessing

Data preprocessing is very important for successful and reliable models. This study did perform some cleaning on the raw data to remove missing and outlier data. I handled the missing values by using Multiple Imputation to build some complete data sets so I could fill in the missing data points, making the information more full and correct. As for the outlier analysis, the study used 3σ to find the outlying value, which means the mean - 3 standard deviations is an error value, then this data point is deleted or corrected.

After cleaning up the data. I applied One-Hot encoding to the variable. One - Hot Encoding is the way we change categorical variables into binary vector so that model can process those variables. Like, the administrative experience variable became into multiple binary variables which stood for

a certain kind of post. When it comes to continuous variables, we apply standardization and turn the data into a distribution with a mean of 0 and a standard deviation of 1, in order to get rid of dimensional differences among variables and guarantee the stable training of models.

Once data preprocessing was complete, the dataset was split into training and testing sets in a 6:4 ratio. The training set was used for model training and parameter adjustment, while the testing set was used to evaluate the model's generalization ability. This division helps to effectively avoid overfitting and ensures that the model performs well on new data.

3.2. Model construction and training

During the model training phase, this study used the XGBoost regression algorithm. XGBoost (eXtreme Gradient Boosting) is an algorithm based on the gradient boosting framework, where the core idea is to optimize the model by iteratively adding new decision trees [3]. In each iteration, XGBoost calculates the gradient of the loss function of the current model with respect to each feature and constructs new decision trees based on these gradients. XGBoost fully leverages the interactions between features, automatically selects and combines features, thus improving the accuracy of model predictions. The relevant computation formulas for XGBoost are shown in equations (1)-(4).

$$L(\emptyset) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (1)$$

$$[L^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T] \quad (2)$$

$$[G_j = \sum_{i \in I_j} g_i] \quad (3)$$

$$[H_j = \sum_{i \in I_j} h_i] \quad (4)$$

4. Findings

This document conducts an in-depth analysis of national outstanding teacher data (from the Beijing-Tianjin-Hebei region) based on the XGBoost regression model, aiming to explore the key factors influencing the academic output of outstanding teachers. The model uses “number of papers published” as the prediction target and selects seven features, including “administrative experience,” “expert training,” “number of awards,” etc., as independent variables. The XGBoost regression model was established using the training dataset and applied to the testing data for prediction.

Through feature importance analysis with the XGBoost model, the study found that “expert training” has the greatest impact on teachers’ academic output, accounting for 50.20%. The next most significant factor is the “number of awards,” which accounts for 30.80%, followed by “administrative experience,” which accounts for 17.50%. On the other hand, “collaborative ability,” “overseas study and exchange experience,” “excellent teaching examples from the Teaching and Research Network,” and “practical experience” have a smaller impact on academic output, with respective proportions of 0.00%, 0.60%, 0.90%, and 0.00%. These results suggest that teachers’ training experiences and award achievements are the key factors influencing their academic output, while administrative experience also has a certain impact.

Table 1. Feature importance

Feature Name	Feature Importance
Administrative Experience	17.50%
Expert Training	50.20%
Number of Awards	30.80%
Collaborative Ability	0.00%

Table 1. (continued)

Overseas Study and Exchange Experience	0.60%
Excellent Teaching Examples (Teaching and Research Network)	0.90%
Practical Experience	0.00%

It shows some predictive ability for the training set with an MSE of 6970.217, an RMSE of 83.488, and an MAE of 40.938. These metrics show the model's fitting on the training data. The MAPE of 1663.908 and the R^2 of -0.296 indicate that the model's performance on the testing set is unsatisfactory, but the results on the training set still serve as a basis for reference when optimizing later. These data show that the model has developed some degree of learning capability and has laid a solid foundation for improving its performance. By modifying the model structure or optimizing parameters, we hope to improve the generalization capability of the model.

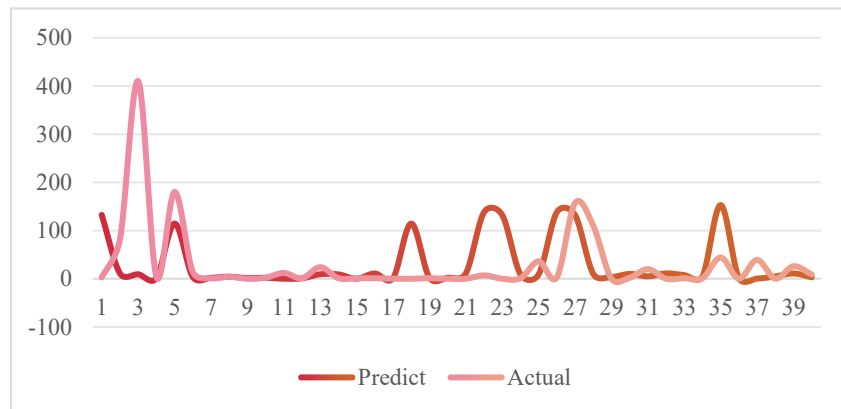


Figure 1. Model prediction performance

5. Discussion

Academic output of teachers who were once expert trainers was larger, and this phenomenon can be explained from several angles. First, the role of teachers as training experts requires them to constantly update and enrich their own knowledge and information so that they can teach others better. This continual learning process both sharpens teachers' competence and gives new thoughts and inspiration for their school-based research. Second, in the course of training, teachers come into contact with all kinds of people: This sort of exchange is good at broadening their horizons, promoting new thoughts, and doing deeper, broader academic study.

Number of award is an important indicator of teachers' academic achievement and also an important influence on the teacher's academic results. Rewards are a kind of high-level academic recognition to teachers. It can improve the academic confidence of teachers, and also encourage them to achieve higher academic goals. Teachers tend to be rewarded, which usually improves the academic status and scope of influence. It becomes convenient for them to get access to research resources, chances of collaboration, and further academic support. This sets favorable conditions for their future research efforts.

Administrative positions need teachers to see things from a broad perspective and strategically. This allows them to be kept with the latest trend on the current day's academic study and plan out for the long run [4]: Therefore, even though administrative experience cannot directly determine academic output, it can indirectly make teachers better so that they can be more successful academically.

6. Conclusions and educational management suggestions

Teachers' professional development is a multi-faceted process, experiences of being training experts, academic honors, and administrative positions are important too. Schools and educational institutions ought to guide teachers through policies and mechanisms and development of culture so they can appreciate all the roles and have the appropriate support and safeguard for them. Also, teachers themselves should actively embrace many roles and improve their various abilities as well in order to have a breakthrough in their professional growth.

Firstly, schools and education institutions need to make teachers realize the importance of such a role and the role of experts in training academic ability. Schools can organize special lectures, experience exchange meetings and other activities to promote the teacher's participation in training work so that it can be carried out in both words and deeds. [5]

The second thing, schools must intensify the promotion and explanation of academics honors, so teachers could know how awards are significant for individuals' own career growth and academic status. schools can create a school atmosphere of cherishing learning and striving for excellence by organizing activities like award-winning teacher experience sharing activities and academic achievement shows. We also need to establish an all-round academic reward system. Academic rewards include various material and spiritual rewards for the teaching staff and faculty who receive good results in scientific research. This can promote more people to continue their new innovations.

Third, schools should guide teachers to correctly understand the meaning of administrative work and motivate promising teachers who wish to take on administrative responsibilities. Schools give some training and support so teachers get better at organizing and running stuff, finding resources, and managing teams — this means the teachers have lots of help when it comes to doing their schoolwork.

References

- [1] Imani, R., Simonds, R., Sharps, D., Wallace, J., Joseph, H., Ropac, R., & Bergin, T. (2025). Integration of individuals with lived experience to improve recruitment within criminal justice research: 'Experience as the best teacher'. *International Journal of Social Research Methodology*, 1(1), 1–14.
- [2] Ignacio, J. A. G. (2024). Exploring the perspectives of preservice Filipino mathematics teachers: Basis for a proposed educational belief model. *Problems of Education in the 21st Century*, 4, 487–506.
- [3] Panchal, P., Shingala, B., Thakor, S., Jain, P., Vaja, C. R., Joshi, A., ... & Pathak, J. (2025). XGBoost regression analysis of dielectric properties of epoxy resin with inorganic hybrid nanofillers. *Journal of Macromolecular Science, Part B*, 3, 347–363.
- [4] Hallinger, P., & Kovačević, J. (2022). Mapping the intellectual lineage of educational management, administration, and leadership, 1972–2020. *Educational Management Administration & Leadership*, 2, 192–216.
- [5] Bağlibel, M., & Arslan, M. C. (2014). Alumni views about educational administration, supervision, planning, and economics non-thesis master's program: The case of Gaziantep University. *Gaziantep University Journal of Social Sciences*, 4, 1079–1095.