



# Analyzing Student Achievement: Data Processing Models in Education

Lei Wang

Ningbo foreign language school, China

\*rain\_wanglei@outlook.com

**Abstract.** Multiple factors contribute in a non-linear manner, making the field more attractive. The wide availability of educational datasets further fuels this interest. The potential application of data processing models in the field of education is: Data analytics help teachers to understand the learning abilities and challenges of their students and promote a deeply ingrained cultural process of using detailed inputs (information) to ensure optimal outputs (student outcomes). The purpose of this article is in this literature it will show the prediction of student achievement. First of all, it is necessary to analyze the data so this requires some Python models like Pandas, also some charts like pie charts, bar charts and so on. With the increasing size of schools in our society, the problem of ensuring and improving the quality of teaching is becoming more and more prominent, and various teaching research and teaching practices are emerging. Evaluation is used to improve the quality of teaching and to motivate students to study hard.

**Keywords:** Analyzing, Predicting, Chart, Pandas.

## 1 Introduction

Academic achievement is the most important and basic index to evaluate students' learning status. The analysis of learning performance is helpful for teachers to grasp students' learning situation and give targeted instruction. Student behavior data is a reflection of students' learning style and living habits at school. Students can also find and develop their own study plan according to their own academic performance.

Achievement analysis refers to the quantitative and qualitative analysis of a student's academic performance in order to identify the strengths and weaknesses of the student in the learning process.

Clear learning objectives: Through the analysis of grades, students' learning level and needs can be understood, and appropriate teaching strategies can be formulated to achieve the learning objectives of the class as a whole and students individually;

Optimize the teaching process: Performance analysis can often reflect the bottlenecks and problems in the teaching process, and through improvement, it can provide students with a better learning experience [1-5].

Improve the quality of teaching: teaching problems, learning weaknesses indicate the shortcomings of the management system or quality risks, identify and improve, can gradually improve the quality of teaching.

In modern society, students are susceptible to many factors that interfere with their education and career. With this social factor in mind, the Student Achievement Analysis and Advisory System (SPACS) system will play a crucial role in analyzing the overall academic performance of students. It identifies factors that affect performance, helps advise students to improve performance, and provides guidance on choosing a career path by providing assessment reports. The three main aspects that determine the right career path for a candidate are divided into entrepreneurship, research and employability.

Application of Data Processing Models in Analyzing Student Performance in the Field of Education.

Teachers have the ability to interpret students' work and determine their level of mastery based on learning standards. However, constantly determining student performance is easier said than done. In order to effectively measure student progress, teachers need to examine students' daily work and growth over the course of a year. This means delving into data big and small-test scores, student grades, homework completion rates, and more.

Research on the impact of different factors on student achievement.

Uncomfortable learning environment, family background, learning infrastructure, difficulty in understanding, teacher-student ratio, information overload, pressure to perform, unhealthy lifestyle, distraction and loss of interest are the factors affecting student achievement [6].

## 2 Methodology

In this project, it is best to use pandas to handle the analysis. The panda is a kind of python. Pandas' main data structures are one-dimensional data and two-dimensional data, which are sufficient to handle most typical use cases in finance, statistics, social sciences, engineering, and more. For R users, Data Frame provides richer functionality than the R language data frame. Pandas is based on NumPy and integrates perfectly with other third-party scientific computing support libraries.

We can use pandas to:

Read the data in CSV, Excel and so on.

Statistics

Clear data

Draw bar charts, histograms, or other visual graphics with the help of Matplotlib

Store the processed data in an Excel or CSV file

Introduction of Pandas Library

Pandas is an open source library in Python that is primarily used to easily and intuitively work with relational or labeled data. It provides various data structures and operations to manipulate numerical data and time series. This library is built on top of Python's NumPy library. Pandas is fast and offers high performance and efficiency to its users. Pandas is Python's core data analysis support library, providing fast, flexible, and unambiguous data structures designed for simple and intuitive handling of relational and labeled data.

pandas in order to be a necessary high-level tool for Python data analysis practice and practice, with the long-term goal of being the most powerful and flexible open source data analysis tool that can support any language. After years of unremitting efforts, pandas already get closer and closer to this goal.

This type of data is used to assess student performance.

Student learning data are quantitative, numerical data from grades and test scores, as well as a qualitative, corrective notes, field notes, and teacher feedback. Most schools have systems in place to collect and assess student learning data.

Using Student Performance Data to Improve Student Achievement

Teachers can use student data analytics to identify factors that may motivate student performance and then adjust their instruction to better meet student needs. Explain expectations and assessment standards.

Of all the courses children take in school, math performance is of particular interest to many parents because it shows a child's ability to understand, compute, apply, reason, and participate. These five characteristics are also interdependent. We are also interested in math achievement because students all over the world take math courses, and we believe that studying math achievement is more representative and meaningful in understanding student performance [7-10].

### 3 Experiment and result analysis

**Table1.** Table student performance.

| gender | race/ethnicity | parental level of education | lunch        | test preparation course | math score | reading score | writing score |
|--------|----------------|-----------------------------|--------------|-------------------------|------------|---------------|---------------|
| female | group B        | bachelor's degree           | standard     | none                    | 72         | 72            | 74            |
| female | group C        | some college                | standard     | completed               | 69         | 90            | 88            |
| female | group B        | master's degree             | standard     | none                    | 90         | 95            | 93            |
| male   | group A        | associate's degree          | free/reduced | none                    | 47         | 57            | 44            |
| male   | group C        | some college                | standard     | none                    | 76         | 78            | 75            |
| female | group B        | associate's degree          | standard     | none                    | 71         | 83            | 78            |
| female | group B        | some college                | standard     | completed               | 88         | 95            | 92            |
| male   | group B        | some college                | free/reduced | none                    | 40         | 43            | 39            |
| male   | group D        | high school                 | free/reduced | completed               | 64         | 64            | 67            |
| female | group B        | high school                 | free/reduced | none                    | 38         | 60            | 50            |
| male   | group C        | associate's degree          | standard     | none                    | 58         | 54            | 52            |

|       |         |                    |          |      |    |    |    |
|-------|---------|--------------------|----------|------|----|----|----|
| male  | group D | associate's degree | standard | none | 40 | 52 | 43 |
| femal | group B | high school        | standard | none | 65 | 81 | 73 |
| e     |         |                    |          |      |    |    |    |

---

This is done by using data that is only part of the display. It includes the student's gender, their group, their parents' level of education, and scores in different subjects. I can use these scores to calculate the student's total score to achieve my purpose.

At the same time, it also shows that there are many factors affecting the score. The main reasons: such as talent, health, self-discipline, effort (preparation, practice), learning methods, memory ability, goals, exam skills, psychological quality (emotional management ability, focus, willingness, stress resistance, confidence, etc.), time management ability, learning environment (desk mates, classmates) tutor, exam difficulty, extracurricular reading, region, etc.

Secondary reasons: such as parents' education level, family economic status, siblings' academic performance, living environment, friends, eating habits, entertainment hobbies, etc.

As shown in [Table 1].

All in all, these data can help me predict the total score because of this information. This reduces the standard deviation. Computers can make more accurate predictions.

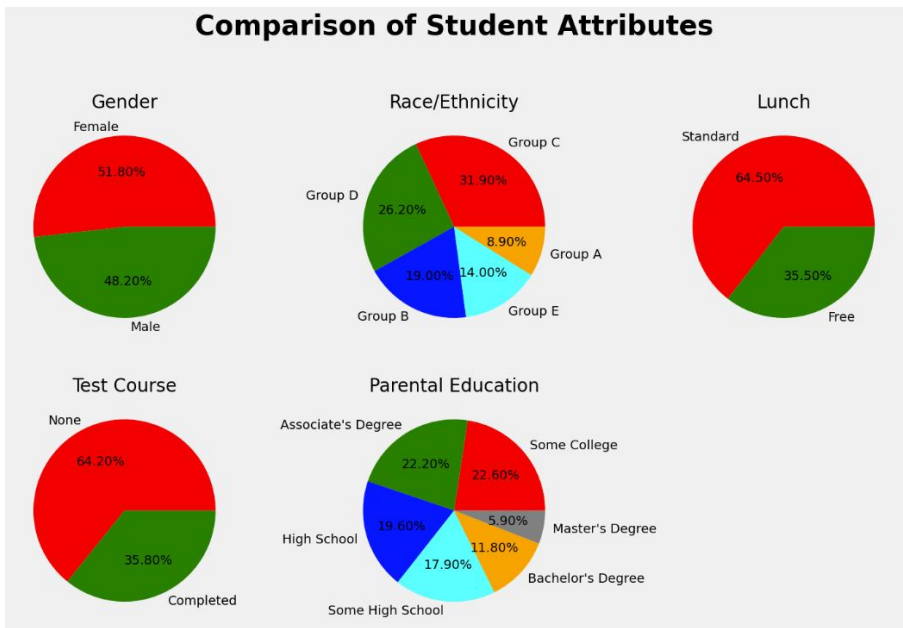
Parents and educators are always concerned about children's performance in school, because school performance is related to the long-term development of children. A child's grades in school are an important aspect of showing how a child is doing because grades are based on how much effort a student puts in and their willingness to learn new things. Besides, a good grade can motivate children to build up confidence. Therefore, a child's grade is an important aspect of the study of child performance.

Of all the subjects children study in school, math scores are of particular concern to many parents because they can show a child's ability to understand, calculate, apply, reason, and engage. These five characteristics are also interdependent. We are also interested in math scores because students all over the world take math courses, and we feel that studying math scores is more representative and meaningful to understand student performance.

Why is it important for teachers and students to analyze test data

Teachers can have a deep understanding of students through performance analysis, so that they can better help students, such as improving the areas where students are weaker.

Parents and educators are always concerned about children's performance in school, because school performance is related to the long-term development of children. A child's grades in school are an important aspect of showing how a child is doing because grades are based on how much effort a student puts in and their willingness to learn new things. Besides, a good grade can motivate children to build up confidence. Therefore, a child's grade is an important aspect of the study of child performance.



**Fig. 1.** The pie chart of the factor can affect the scores (Photo/Picture credit :Original) It includes five factors can impact the student performance. And the proportion of these factors.

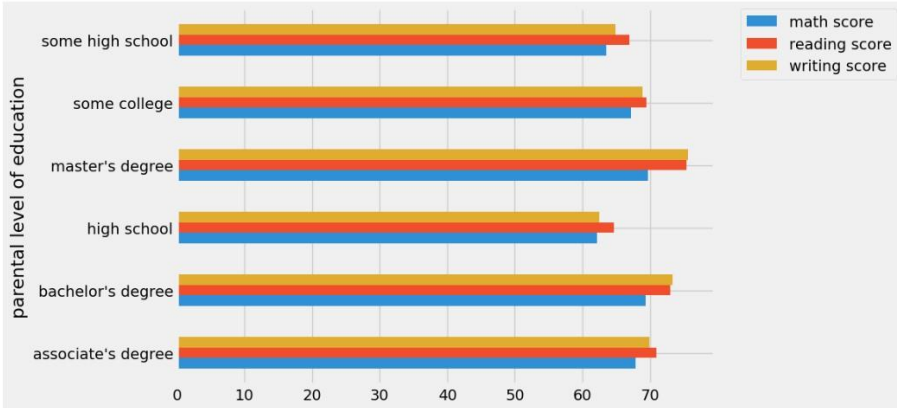


Fig. 2. The bar chart about the parental level education (Photo/Picture credit : Original)

It includes six parental level of educations and the three scores in different parental level of education.

Using pie chart and bar chart to display the data from the form.

For this experimental procedure, we first define the problem statement and investigate the algorithm and regression implementation pipeline that we will use. We then go on to actually implement recognition and regression algorithms such as linear regression, Lasso, K-nearest neighbor regression, decision trees, random forest regression, XGB regression, Cat Boosting regression, and AdaBoost regression. Next, we compared the performance of these models. Finally, we build a linear regression model and prove that it is most suitable for the student achievement prediction problem. We can use python. First, we need to make a table of student achievement data in python. Second, use Python to check if there are any problems in the table, such as whether there are any losses. Then the data is converted from classification to numerical value, making the analysis of the data more convenient for the computer. Finally, we can use standard deviation to show the result of the prediction. It is important to analyze the student's progress. Different disciplines have different assessment criteria, some are subject specific and some are common across disciplines. On the other hand, learning styles and learning patterns also play an important role in students' performance and progress in different assessment criteria. We have developed student attribute descriptors to provide a more complete picture of student progress and performance.

This analysis tells us that factors such as parents' education level, socioeconomic disadvantage, and test preparation courses all affect students' test scores. But there are many exceptions. Some students whose parents have low education levels get full

marks. In addition, some students do not complete exam preparation courses and get full marks. These students may have their own test preparation strategies. There are many exceptions to socioeconomic disadvantage. These students did not let financial barriers hinder their efforts. So, there are many factors that affect student performance. Some have a big impact, and some don't. In addition, there are other factors to consider that are not mentioned in the dataset. Factors such as school facilities, teaching quality and methods, peer pressure, learning time, diet, sleep patterns, etc. These factors also affect student performance.

In addition, the model shows that students' math scores are highly correlated with the school we use, gender, age, parents' education level, length of study, number of past course failures, additional educational support, home educational support, etc., so many variables are one of the reasons for our high accuracy [Figure 1]. In real life, many things affect student achievement in school, which is the same as what we analyzed in the project. Therefore, if a person wants to improve his/her performance in school, it is not an easy thing, because he/she needs to put more effort in his/her study and change his/her study habits. While the results were somewhat encouraging, there were some serious problems with the dataset. First, there were problems with the sampling procedure used to collect the data[11].

## 4 Conclusion

Predicting student achievement can help them be proactive and steer students toward areas that need improvement. Academic achievement is the most important and basic index to evaluate students' learning status. The analysis of learning performance is helpful for teachers to grasp students' learning situation and give targeted instruction. Student behavior data is a reflection of students' learning style and living habits at school. Based on the behavioral data of college students, this paper uses data mining technology to fully mine and analyze the life habits of college students, studies the potential behavioral patterns with strong correlation between student behavioral data and student grades, and then establishes a grade prediction model through data mining technology. To achieve the purpose of using students' daily behavior to predict students' academic performance, establish student counseling mechanism and improve students' academic performance.

At present, the development and utilization of educational big data in China has already had a certain foundation, but there are still problems in the actual business



such as insufficient amount of data, insufficient data source, single prediction method and low prediction accuracy. These problems are the bottleneck that education big data needs to solve.

When teachers are forced to examine this data by hand, it quickly leads to burnout. Entering each assignment and course grade individually is cumbersome and time consuming. We both know you already have a lot on your plate. Fortunately, advanced technology designed for schools makes it easier than ever to measure, manage, and evaluate student performance data.

By analyzing student performance data, you can spot anomalies that point to gaps in understanding. For example, let's say you've just finished grading your first exam of the year, and you notice that a group of students don't understand a certain concept. In this case, you can use the performance data to backtrack and determine where the loss of comprehension occurred. You can then develop a differentiated lesson plan to ensure that the group gets back on track without disrupting the rest of the class. In the end, it was very friendly for us.

## Reference

1. Ahmedameenhassan, AHMED AMEEN HASSAN. "Students Performance." Kaggle, 21Feb. (2022)
2. [www.kaggle.com/code/ahmedameenhassan/studentsperformance/input](https://www.kaggle.com/code/ahmedameenhassan/studentsperformance/input) . (2023)
3. Team, PythonGeeks. "Introduction to Python Pandas." Python Geeks, 1 Oct. 2021,
4. [pythongeeks.org/introduction-to-python-pandas/](https://pythongeeks.org/introduction-to-python-pandas/). (2023)
5. Staff, EasyBib. "EasyBib®: Free Bibliography Generator - MLA, APA, Chicago Citation Styles." EasyBib, Chegg, 1 Jan. 2023, [www.easybib.com/](https://www.easybib.com/). (2023)
6. Khan, Anupam, and Soumya K. Ghosh. "Student Performance Analysis and Prediction in Classroom Learning: A Review of Educational Data Mining Studies - Education and Information Technologies." SpringerLink, Springer US, 1 July (2020) [link.springer.com/article/10.1007/s10639-020-10230-3](https://link.springer.com/article/10.1007/s10639-020-10230-3). (2023)
7. Jurgens, Tyson. "Evaluating and Optimizing Student Performance with Data." Schoolytics,Schoolytics, 4 July (2023)
8. [www.schoolytics.com/blog/2022/09/evaluating-and-optimizing-student-performance-with-data](https://www.schoolytics.com/blog/2022/09/evaluating-and-optimizing-student-performance-with-data). (2023)
9. "Sign In." RPubS, [rpubs.com/stellazhong929/860612](https://rpubs.com/stellazhong929/860612). Accessed 25 Aug. (2023).
10. IEEE Xplore, [ieeexplore.ieee.org/Xplore/guesthome.jsp](https://ieeexplore.ieee.org/Xplore/guesthome.jsp). Accessed 25 Aug. (2023).

- 11.Chavan, Sanket. "Students Performance Analysis." Medium, Analytics Vidhya, 28 Feb. 2021, [medium.com/analytics-vidhya/students-performance-analysis-46eba510039c](https://medium.com/analytics-vidhya/students-performance-analysis-46eba510039c). (2023)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

