Some crucial aspects differentiate state of the art for learning performance prediction research. Existing research on Learning Performance Prediction can be divided according to data(input), predicted values(output), and assumption.

1. **Input data:**

Performance prediction has different prediction effects because of the different input data.

The input data in the existing research can be divided into direct factors and indirect factors.

* 1. **Direct factors:**

Direct factors refer to the original data in the education system. These data can be divided according to the type of education system.

**In Traditional education systems:**

The commonly used mining data in the traditional education system includes course performance, demographic feature, and learning material.

1. **Learning outcomes:**

Existing articles usually use pre-course performance data or current course performance, assuming that the situation in the past or the current stage can be used to predict the future.

In the traditional education system, these course performance data can be quiz results, test scores, accuracy rate, GPA for a semester, a school year, graduation, rewards and punishments, and homework completion.

1. **Demographic feature:**

Demographic data is commonly used in traditional education systems, including family, Gender, Age, Level of Schooling, Country of Origin, Primary Language, Employment Status, etc.

**In distance education systems:**

distance education system refers to a learning system for teaching and learning online. Compared with traditional education systems, distance education systems have more complex and diverse data types, and data records are complete and coherent so that more information can be obtained from them. The unique data types in the distance education system are learning material and learning behaviour data. Learning behaviour data refers to the data that students interact with the system when they participate in online teaching.

1. **Learning material:**

There are many learning material types in the distance education system, such as teaching videos, text, and picture courseware.

1. **Learning behaviour data:**

learning behaviour data refers to the log and engagement data generated by the interaction between students and the system during online learning. Here, learning behaviour data is divided into video related, exercise-related, forum related and platform use according to online teaching.

* **Video-related:** learning behaviour data contains data generated by student interaction during video learning and video software teaching. For example, log data such as play, pause, skip, adjusting rate, etc. when watching the video, and students answering questions and discussing during the video lecture.
* **Exercise-related:** exercise-related learning behaviour data refers to related data such as quiz, assignment, and test. For example, test scores, whether the exercises are CFA (Correct at First Attempt), Distinct Problems Attempted, Submissions, Time Spend on One Problem, Number of Correct Problems, etc.
* **Forum related:** contains data like Negative/Positive Rating Number, Posts Number, Words Number
* **platform use:** platform use data is platform-related data in addition to other data. Contains such as the number of access, the period of attending courses, access time, and registration time, etc.
  1. **Indirect factors:**

Indirect factors refer to factors that require processing to be indirectly extracted.

1. **Social network**

[1]–[4]

1. **Personality**
2. **Learning style**
3. **Predicted values**
   1. **Learning performance:**

Commonly used outputs to measure student learning performance prediction include GPA, CFA (Correct on First Attempt), Exam / Test Score, Course Grade Range, Program or Module Graduation / Retention / Dropout, Assignment Performance, Number of Passed Course, Certificate Earners, etc.

* 1. **Other predicted values:**

1. **Interaction Between Learner**
2. **Concept Associations**
3. **Question Difficulties.**
4. **Student Behavior**
5. **Assumption**
   1. **Sequential**
   2. **Non-sequential**