Prediction Part

Basic concept: The main purpose of this prediction part is to make reasonable prognosis automatically. As we can image a season teacher can make an accurate guess of a particular student based on his or her current study, the aim is to simulate those seasoned teachers to help predict for larger scale of student automatically, which can save the time for human teacher to absorb data and analyze manually and also enhance the prediction accuracy.

Input:

\*\*Behavioral data: all the data except the direct result (all type of grade) For example, the tutorial attendance in traditional data and learning progress log in online data. Intuitively human teachers make reasonable guess based on those data so we assign much importance to those behavioral data closely related to the learning process. Also we use some data like social platform surrounded the learning platform to try to figure out more reliable indicators to help us enhance accuracy. （介绍data是啥，简单说下为啥用来predict）

Output: expected result -> final grade, we map all results to five classes, A, B, C, D, F, with grade decreasing one after another. In another word, this part actually classify different learning styles to different grade class. It could be more tolerated compared to the model directly predict the result in percentage system and more diverse compared to binary classification. （介绍下输出结果的形式，现在暂时认为这样之后根据implement改，说说这样输出的好处如果有的话）

Analysis method –machine learning:

We use both traditional statistical learning and deep learning in our analysis.

Because we are not familiar to the real learning analytics data, we are trying to pick several popular analysis models which can handle linearly separable data and other data with more complicated nature.

We adopt a powerful python machine learning library to facility us applying statistical learning methods. And we adopt the library function to auto select the best model with highest accuracy to process the data. In essence, we separate the analysis part as a black box, so that we can focus on the server building and event handling.

As for deep learning part, we adopt another python library to test our assumptions to use ANN to make some difficult prediction such as predict final grade at the beginning of the course.（不一定，只是说说需要讨论一下神经网络的使用范围和优势）

Advantage:

1. At the very beginning we decided to use machine learning strategy to tackle the prediction part because we think the process to make a prediction in this case is almost the same as a skillful and experienced teacher make predictions by previous experience.
2. It’s more easy to use machine learning library rather than developing our own judgement strategy to pick up the student we need, and the accuracy can be more stable with more training data.

Other functionalities based on the prediction model: (1) can filter out top student study habit model

(2) define risk factor-> can find the student with high risk and make alert

(3) comparison different student study model to give corresponding (predifined) result

(4) can figure out some suspect students with unmatched study style and current result（新想出来的，有可能找出作弊的人）

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*something jot down\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

One difficult point of collecting data:

Not all-rounded.

Some data like social platform will be hard to collect and involve some privacy issues.

Data preprocessing: give 极端情况 larger weight training to simulate some prominent cases in the experience of a teacher