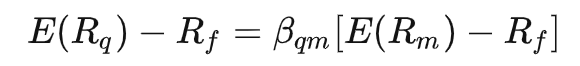
1. **Introduction**

In this project, we selected 6 stocks listed on Shanghai or Shenzhen stock exchange and applied machine learning methods to conduct asset pricing and developed several ML backed return prediction models (next trading day’s return).

Our stock selection criterion is based on traditional asset pricing model – CAPM. CAPM is a very widely used model which creates link between the risk and expected return of stocks. Since outer market environment is hard-to-predict, we should balance our investment through balancing composition risks. Therefore, we selected 3 stocks with higher-than-market risks, namely SZ002415, SZ002475, and SH601012, and 3 on-the-contrary stocks, namely SH600009, SH601390, and SH601668. The ratio of radical growing and steady growing stocks is 1:1 so as to empower the effect of risk hedging when market expectations are unpredictable.

1. **Stock selection detail**

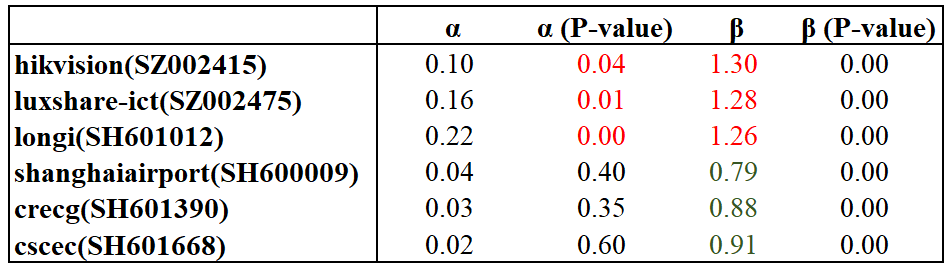
* **Method**



We utilized CAPM model to select stocks. Where E(Rq) denotes expected return of stocks; Rf​ denotes risk-free rate (we use weighted average of 5-year treasury bonds and 5-year time deposits as Rf); βi​ is beta value of the investment that is a measure of how much risk the investment will add to a portfolio that looks like the market.

* Result​

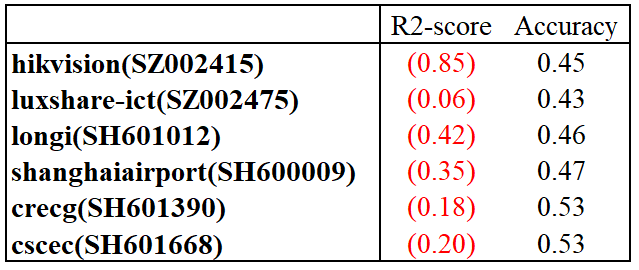
After applying CAPM to more than 50 stocks, we selected 6 stocks and computed their α and β value:



With β value significantly larger than 1, a stock can be regarded as a radically growing stock with greater risk than investing on market index. And when α value is significantly larger than 0, a stock’s overall trend would be significantly stronger than the market.

1. Baseline prediction

Since CAPM could be used in stock pricing, we then used CAPM to conduct future return prediction. We used lag-1 term Rm and Rf to perform prediction on next day stock return, and also constructed a decision index:



The result shows that all predictions have a negative R2, indicating that the baseline model is not efficient, and that the accuracy is round 0.5, also indicating that this prediction performs like a random game.