

---

# Option A

## AI in Finance Final Report

---

**Li Zhuxuan    He Yuhang**  
Department of Mathematics, School of Science  
The Hong Kong University of Science and Technology  
Clear Water Bay, Hong Kong  
{zlidw, yhebj}@connect.ust.hk

### 1 Progress and Learning from Group Project

*Analysis on Codecademy - A company focusing on providing online AI related education*

#### 1.1 Incentives and Purpose

Online programming and AI education companies are of value as AI industry needs talents who are excel at computer science and have comprehensive knowledge about AI. Many online programming education companies provide knowledge and skills related with AI.

The project chooses Codecademy as the analysis objective. This is an online education company focuses on programming and AI education. The reason why choosing this company is that authors consider there is still some investment space according to its financing data.

The purpose of this project is to find out whether this company is worthy investing and learn more about AI industry and future development of AI education.

#### 1.2 Brief Introduction

Codecademy is a web- and mobile-based platform that mainly provides users with programming courses. As AI industry needs talents who are excel at programming, especially python, codecademy can provide this kind of online education. The article will mainly focus on the competitor analysis from four different aspects: product , price, promotion and financial data analysis.

After research, we choose three competitors of Codecademy: w3schools, Code School and Coursera.

#### 1.3 Product Analysis

For the product part, the core is the learning contents and the designing and convenience of the website.

##### 1.3.1 Learning Contents

Codecademy provides free and non-free programming courses for users. Its languages are rich, including python, C++, HTML, JavaScript, PHP, Ruby. The best course it can provide is Python, which is very related to AI and finance. Users can easily use it when learning programming as the learning interface is divided into three columns: the first column is the guide, the second column is the input area and the third column shows the result. Target users of Codecademy are people who want

to learn systematic programming and do not learn about programming before, and are interested in AI.

W3schools is an online platform that provides its users with education related to web development. It provides its users with tutorials related to HTML, HTML5, CSS, CSS3, Java, JavaScript, jQuery, jQueryMobile, AngularJS, AJAX, JSON, SQL, PHP, ASP, ASP.NET, VBScript, AppML, XML Tutorials, WEB Building, and more. Target users of w3school are people who want to learn programming and mainly focus on web developer and those who have patience to learn on themselves.

Code School provides programming courses for users. Different from other courses, these courses are famous for practical use and are very related to AI. It provides courses combining AI and practical problems, which is very different from other companies. Target users of Code School are people who want to learn programming and AI related knowledge, especially those who need to use programming and AI knowledge in practice and finance.

Coursera provides free (mainly) and non-free courses in many fields (not only programming) for users and these courses are provided by famous universities such as MIT, Harvard. In programming field, it provides many languages: HTML, CSS, JavaScript, PHP, SQL, Python, jQuery, Java. The special characteristic of it is that the courses are very rich and with high quality. Target users are people who want to learn all courses including programming and AI.

### 1.3.2 Designing and Convenience

The design of Codecademy is simple, clear and beautiful, users can easily sign up or look through the price. However, it takes some time to find some information about courses.

W3schools has very simple designing, users can easily find language courses on the left. There are some programming examples on the web page. It is not beautiful and filled of advertisements.

Code School's design is cool, users can easily look through all services it provide via the category. The whole web design is black which gives users a professional feel.

Coursera uses simple, clear designing, users can easily look through all services it provides and the course information.

## 1.4 Price Analysis

| Codecademy                                                                                                                                           | w3school                                                                     | Code School                                                   | Coursera                                                                                                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Price of membership: \$15 – 20 per month, relatively cheap. Members have unlimited access to Codecademy course and paths, and other premium features | All tutorials are free, users can get useful resources (references) for free | Price of membership: \$25 – 40 per month, a little expensive. | Most courses are free if users choose audience mode. If users need a certification, it charges some money. |

Codecademy charges a relatively low price for course membership compared with other non-free online education companies, which indicates that it is more likely to be attractive to users.

## 1.5 Promotion Ability Analysis

### 1.5.1 Company managers and employees

Compared with its competitors, Codecademy's managers are of better education background. Codecademy's founders both graduated from Columbia University. They are very young which indicates that they have passion and energy to develop the company in the future. However, they may not be very experienced to run a company.

Most employees in Codecademy are from excellent schools, such as Harvard, MIT, etc.

### 1.5.2 Number of Visitors and Members

The amount of daily or monthly visitors of these online education platform is important to analyze the investment of these platforms as these data can provide the information of 1) the marketplace of the company; 2) the trends of these numbers indicate the future development of the platform; 3) the future expanding ability of the company.

However, the exact data need to be collected from due diligence of these companies as they are not available for public.

## 1.6 Financial Analysis

Financial analysis is essential to an investment analysis. As Codecademy is a pretty new and growing company, there is not many financial data. However, we can judge the investment value from its financing history.

Codecademy has already raised \$47.5M to develop its operation. The last transaction are supported by Index Ventures, Flybridge Capital, Union Square Ventures, Naspers and Virgin Group and is known as series C. As the capital raised are limited and the financing condition is still in an relatively early stage compared with other competitors, there is still some investment space. Code school was

| Announced Date | Transaction Name  | Investors                                                                                              | Money Raised |
|----------------|-------------------|--------------------------------------------------------------------------------------------------------|--------------|
| Jul 12, 2016   | Series C          | Index Ventures, Flybridge Capital, Union Square Ventures, Naspers, Virgin Group                        | \$30M        |
| Apr 25, 2014   | Convertible Notes | Flybridge Capital                                                                                      | \$5M         |
| Jun 18, 2012   | Series B          | Kleiner Perkins, Index Ventures, Union Square Ventures, Alexis Ohanian, Virgin Group                   | \$10M        |
| Oct 27, 2011   | Series A          | Y Combinator, SV Angel, Slow Ventures, Founder Collective, Union Square Ventures and other 4 companies | \$2.5M       |

acquired by Pluralsight on Jan 26, 2015 at \$36M. This event can prove that this kind of company are very popular in the market.

Coursera has already completed series E financing. It is too late and costly to invest in Coursera now.

## 1.7 Conclusion

More detailed information about Codecademy need to be collected by due diligence to analyze it before investing. From the information collected, Codecademy worth investing for 1) online education mode will be the future trend because of its convenience and effectiveness; 2) data science including

programming will keep popular in the near future together with AI; 3) Codecademy itself provides effective and cost-efficient learning materials; 4) Codecademy still has the need and possibility to continue financing; 5) Codecademy still has the possibility to expand in the future.

## **2 Reflection on An AI Article and A Finance One**

### **2.1 Reflection on the Finance Article (Siliang Ma, & Fuyi Wei. 2018)**

Here want to talk about an article on Southern Finance. This article is quite deep in some aspects and the main contents of this article are as follows. The application of artificial intelligence technology in the financial field has brought tremendous changes to the entire financial industry. Innovative financial services such as intelligent investment, intelligent credit, monitoring and early warning, intelligent customer service have emerged and expanded as the times require. However, the application of AI technology in China's financial field is also facing a series of difficulties and challenges. First, the inadequate basic and key research and development ability of AI technology weakens the core competitiveness of AI technology in the financial field; second, Multidimensional complex data format and huge amount of data bring difficulties to the application, innovation and development of AI technology in the financial field; third, technology security risks increase the difficulty of extensive and in-depth application of artificial intelligence technology in the financial field; fourth, the reform of financial supervision mechanism can not keep pace with the application of AI technology in the financial field. To this end, people should strengthen the construction of talent team of artificial intelligence technology research to enhance the core research and development ability and level; strengthen the integration and utilization of large data resources to build and improve a safe, controllable and widely shared data ecosphere; establish and improve multiple security risk prevention mechanism to enhance the risk prevention and control level of artificial intelligence technology; reform and improve the financial supervision mechanism, achieve full coverage of the application of artificial intelligence technology in the financial field, and provide favorable conditions for promoting innovative development and standardized development of the financial industry.

After reading this article, everyone may have a deeper understanding and Reflection on AI finance.

First, the emergence of AI finance was not accidental. With the intellectualization and convenience of financial transactions and the explosive growth of financial data, the existing financial analysis and forecasting methods don't have the ability to model and analyze such a large amount of data, so some other analysis methods that can be applied to financial data in the context of large data need to be discovered. Previous artificial intelligence technology, combined with new technologies such as big data and cloud computing, can solve the difficult problem of high-dimensional and complex data analysis, and has broad application prospects in the fields of intelligent finance, big data wind control and big data insurance. With AI technology, people can extract the economic logic behind the big financial data from the deep characteristics of financial data, so as to improve the efficiency of financial services and realize the intellectualization of the financial industry. Therefore, AI finance is the inevitable trend of financial development.

However, nowadays AI couldn't be applied to all fields of finance, since finance itself is a complex theoretical discipline. Maybe statistical methods have been widely used in the field of measurement, and more and more machine learning methods have been used in academic research. But in practical application, AI is mainly used to replace a large number of simple and repetitive work. In some more complex tasks, such as electronic customer service and intelligent customer service, it can only deal with some simple and common problems at present. When encountering complex problems, AI needs to be assisted by people to solve them. Similarly, looking at enterprises, it is difficult to have the space for large-scale AI technology in a short time. Therefore, AI can be used better in some financial models with higher profit margin, better data structure and clear problem definition. For example, the hedge fund industry. In the hedge fund market, AI robotic traders move much faster than human traders. When people do foreign exchange or stock trading online, there are robotic traders who can see the opportunity of earnings instantly and quickly carry out trading operations behind the counterparty. That is to say, in the hedge fund industry that has hit the world's financial

markets, the era of exceptional investors with experience and intuitive judgment has come to an end. And another case is that large investment institutions throw out unusually large orders in the futures market included in the SP index. At this time, AI robots also play a great role in ultra-high-speed trading, accelerating the abnormal plunge in stock prices.

But as the article said, the AI financial industry also faces many risks. For example, as the article mentioned, the late start, huge data volume, technical security risks, and imperfect regulatory mechanisms are all risks. Accordingly, in the current life, face recognition and voiceprint recognition are quite common. Some places can pay by people's faces, and some mobile phones are voiceprint recognition unlocked. But the current AI technology is not developed enough, there are still some loopholes in these aspects of recognition that criminals can take advantage of. They can find someone who has the very similar facial features as the object or imitate the object's face through make-up, so that they can steal money or information, and there is a lot of news about these crime cases. There are also some algorithmic flaws that can bring risks. For example, sometimes, people need to rely on machine learning to obtain investment strategies, but finally they found that the results were unsatisfactory in many cases. Later, it was found that there were programming errors and computer vulnerabilities, which caused losses.

In addition to procedural reasons, there are also omissions or errors in factor selection. Everyone knows that when the factors are missing or wrong, the conclusions are often deviated from the actual situation, which is also the reason why it is difficult to predict the stock trend – there are so many factors affecting the stock trend that it is almost impossible to consider it thoroughly. In addition, in the field of economics, the interpretability of models is also very important. After all, empirical science is difficult to be proved as a theory. Most machine learning models at this stage are facing problems with good results but difficult to explain, so many people will ask why you do like this, or require corresponding explanations for each step of decision-making. Maybe this problem can only be succeeded by AI finance.

What's more, with the development of AI finance, there is a growing demand for comprehensive and learning financial talents in the new human-computer interaction mode. Financial science and technology are a combination of Finance and technology, which has the characteristics of professionalism, complexity and cross-border. Ideal financial science and technology talents should know both finance and technology, only one of which is like "lameness". However, for a long time, the relationship between computer and finance had been relatively weak. It's incredibly difficult for a person with a computer background to develop a data model in the financial field, or for a financial practitioner to develop a computer model.

Finally, the article gives some improvement measures, which are not elaborated here. In a word, it is necessary to speed up the pace of science and technology, so that the AI financial industry can reduce its risks while developing, so as to make AI finance serve life more.

According to more materials, AI Finance developing has three stages: computational intelligence, perceptual intelligence and cognitive intelligence. At present, it is in the second stage - perceptual intelligence, that is, it can perceive and process all kinds of input data, including voice, text and image. For cognitive intelligence, the computer actually does not understand the meaning behind these inputs. Only by realizing cognitive intelligence, can humans say that the computer really possesses human intelligence. Therefore, the current application of AI in the financial field is still relatively limited, but with the further development and maturity of AI technology in the future, there will be more innovation and progress in the financial field.

## **2.2 Reflection on the AI Article (Yaobing Lin. 2015)**

Artificial intelligence is a completely different working mode from traditional computer systems. It can read a large amount of "big data" according to general learning strategies and discover rules, links and insights from it. Therefore, AI can automatically adjust according to new data without

resetting programs.

This article seems to analyze the time series data model - a very important part of the current scientific research field. At present, many studies on prediction, simulation and modeling of time series data are mostly based on direct analysis of a model, which has great limitations. Artificial intelligence models commonly used in time series analysis, such as autoregressive moving average model, artificial neural network model and hidden Markov model, can be compared by simulation when selecting research models. The research shows that the mathematical and physical characteristics of time series data are different when different models are generated, which also provides a basis for selecting models.

Autoregressive moving average model is mainly used to predict and fit the trend of time series data, and its function is to establish the model of stationary time series. Moment method and maximum likelihood estimation method are widely used in the correlation coefficient algorithm of the model; artificial Neural Network (ANN) model has only been studied in recent years, and many kinds of network organization forms have emerged, such as self-organizing network, time feedforward network, recursive network, mapping network and so on. The model has been well applied in signal processing, engineering control and numerical calculation. In addition, it also includes the prediction and modeling of time series data; hidden Markov model has been widely used in speech recognition, text information extraction, time series analysis, network path analysis, etc. The model belongs to a kind of double stochastic process, which includes the stochastic process of generating observation value, the stochastic process of hidden state Markov chain, etc.

Through comparative study, it can be seen that among the data generated by these time series data models, the autoregressive moving average model can generate stable short-term correlation data, and the artificial neural network model can generate long-term correlation data, whose structure determines the stability of the generated data, while hidden Markov model generates long-term data, but it is not stable when it is tested by KPSS. Therefore, when choosing a data model, it is necessary to generate data with consistent characteristics with the model. In order to solve this problem, in the artificial intelligence algorithm of selecting time series data model, data should be input first, and then the stationarity and some other correlations of data should be checked. Finally, the selection of artificial intelligence models can be realized by corresponding commands and operations.

Finally, we will talk about the theory of no free lunch. The theory of no free lunch means that there is no best algorithm, only the most suitable one. There are many models for in-depth learning. We must choose models according to our own problems. For example, CNN is very useful for processing image information, while RNN is very good for processing sequence. So we still need to understand each algorithm clearly, choose the appropriate algorithm.

### **3 Synthesis and Suggestions for Further Study**

- For the project part, as Codecademy is a start-up company, much information can be obtained only by due diligence, which is not applicable for the group. Further efforts are needed to complete a perfect investment analysis.
  - Product part: we need to collect users' feedback.
  - Promotion ability:
    1. Estimate companies' expanding ability
    2. Do researches on companies' strategy and marketing power.
  - Financial analysis: need to investigate companies' income sources and their P&L status.
- For the course further study, we suggest that the practical combination of AI with finance is very necessary. Students can do projects in AI companies and learn more about what is happening in the real AI industry.

## 4 Individual Contribution

The followings are individual distribution to this final report.

**Li Zhuxuan:** Responsible for choosing the project theme and finance article, designed the analysis structure and collected information of the project. Represented team to do the presentation.

**He Yuhang:** Responsible for choosing the project theme and AI article, collected information about the project and wrote most parts of part two. Designed the PPT.

### Acknowledgments

Thanks Anthony and Professor Yao for their fully devotion to the course this semester.

### References

[1] Siliang Ma, & Fuyi Wei. (2018) The Application of Artificial Intelligence Technology in Finance: Main Difficulties and Suggestions. *Southern Finance, The 499<sup>th</sup> phase*.

[2] Yaobing Lin. (2015) Artificial Intelligence Algorithms for Time Series Data Model. *Science and Technology Economic Market, Phase 12*

### Appendix A: YouTube URL

<https://youtu.be/rum4qau7jW8>

### Appendix B: Hidden Markov Model

Let  $Q$  be the set of all possible states and  $V$  be the set of all possible observations.  $Q = q_1, q_2, \dots, q_N, V = v_1, v_2, \dots, v_M$  Among them,  $N$  is the possible number of States and  $M$  is the possible number of observations.  $I$  is a state sequence with length  $T$ , and  $O$  is a corresponding observation sequence.  $I = (i_1, i_2, \dots, i_T), O = (o_1, o_2, \dots, o_T)$ ,  $A$  is a state transition matrix (a state transition matrix at a time):

$$A = [a_{ij}]_{N \times N}, \quad I = 1, 2, \dots, N; \quad J = 1, 2, \dots, N$$

Under the condition that time  $t$  is in the state of  $q_i$ , the probability of time  $t + 1$  transferring to the state  $q_j$  is as follows:

$$a_{ij} = P(i_{t+1} = q_j | i_t = q_i)$$

$B$  is the observation probability matrix (one observation probability matrix at a time):

$$B = [b_{jk}]_{N \times M}, \quad K = 1, 2, \dots, M; \quad J = 1, 2, \dots, N$$

The probability of generating observation  $V_K$  under the condition that time  $t$  is in  $Q_J$  is as follows:

$$b_{jk} = P(o_t = v_k | i_t = q_j)$$

$\pi$  is the initial state probability vector:  $\pi = (\pi_i) \pi = (\pi_i)$ , Where  $\pi_i = P(i_1 = q_i)$  Hidden Markov model:  $\lambda = (A, B, \pi)$ .

### Appendix C: Artificial Neural Network Model

There is a time series:

```
u=[17.6 17.7 17.7 17.7 17.8 17.8 17.9 18.0 18.1 18.2 18.4 18.6 ...  
 18.7 18.9 19.1 19.3 19.6 19.9 20.2 20.6 21.0 21.5 22.0 ]
```

We can use neural networks to predict the next few items. Codes are displayed below:

```
clear  
u=[17.6 17.7 17.7 17.7 17.8 17.8 17.9 18.0 18.1 18.2 18.4 18.6 ...  
 18.7 18.9 19.1 19.3 19.6 19.9 20.2 20.6 21.0 21.5 22.0];  
Want to use neural networks to predict the next few items  
n=length(u);  
for i=1:(n-3)  
    p(i,:)=u(i:i+3);
```

```

if i~=n-3
    t(i)=u(i+4);
else
    t(i)=22.7;
end
end;
net=newff(minmax(p'),[n-3 20 1],{'logsig','logsig','purelin'},'trainlm');
net.trainparam.show=20;
net.trainparam.lr=0.01;
net.trainparam.epochs=300;
net.trainparam.goal=1e-3;
net.trainParam.mc=0.99;
net.trainParam.delt_inc=1.2;
net.trainParam.delt_dec=0.3;
net.trainParam.deltamax=50;
net.trainParam.time=inf;
net.trainParam.mem_reduc=1;
net.performFcn='sse'
net.layers{1}.initFcn = 'trainlm'
net.inputWeights{1,1}.initFcn = 'trainlm'
net.biases{1,1}.initFcn = 'trainlm'
net.inputWeights{2,1}.initFcn = 'trainlm'
net.biases{2,1}.initFcn = 'trainlm'
net=train(net,p',t)
sim(net,[20.6 21.0 21.5 22]')
#22.700

```

We can get the prediction result: 22.7.