

Employing modern Artificial intelligence and Machine Learning strategies, and human trader experience for R&D S&P 500 index forecasting models. (Appendix and Evidence)

Name: Chow Chi Kin

Student No: 199172502

Title: PROM02 - Computing Master's Project (2020/1 - UOSHK - K07 - NSSEM3)

## 1 Project Achievements, Challenges and impacts

### 1.1 Achievements

#### 1.1.1 Conducting a literature review of financial forecasting

Since clients have technical analysis and statistical methods to forecast S&P 500 trends, we evaluated different research methods by literature reviews and want to discover which method is able to reach the project goal within a short project timeframe.

Literature review skill is applied to search for and access publications, reading and analyzing sources, evaluating data and publications to determine which literature makes a noteworthy contribution in financial forecasting.

Based on the literature review, we found that CNN image recognition methods are able to be applied in client's technical analysis to recognize his chart patterns trading strategy[1]. Regression methods are able to be applied in his statistical trading method. [2]

#### 1.1.2 Successful ETL

Client used the statistical trading method to forecast S&P 500 trend and was able to keep away the 2018 falls[3]

We used the window function to transform the raw data to EMA and MACD signals as feature data. Also, we transformed index data to last day percentage change as feature data.

All algorithms are able to gain positive annual revenue return in the 2018 market fall because we successfully transformed the client's statistical trading method as feature data.[figure 1]

#### 1.1.3 Greater than S&P 500 annual return

Client requests using tensorflow to trade S&P 500 and gain annual return greater than S&P 500 annual return.

We used ensemble neural network (using stack method to ensembling the Neural network with mse function and neural network with custom loss function) to forecast the buy and sell signal.

Ensemble Neural network able to win S&P 500 7 year's annual return from 2011-2020. Total return is greater than S&P 500. Client is satisfied with the result. [figure 5]

## 1.2 Challenge

### 1.2.1 Completed TensorFlow tutorial and course

Tensorflow is the most popular machine learning and AI API in the world. Client requests to use it for the project. Since we do not know which machine learning method will be used for the project at the beginning, researchers finished Tensorflow's official tutorial[4] and Udacity's Tensorflow course[figure 2].

Tutorial and course included regression, time series forecasting and neural network. We applied linear regression[5] and neural networks[6][7][8] into the project for the client's statistical trading method.

### 1.2.2 Learnt Python 3.0

Most of the TensorFlow tutorial and example is programmed in Python 3.0. In order to get more resources using tensorflow. Researcher is switched from using R to Python and finished a python course from Datacamp[figure 3]

Course is include Python basics and Numpy. We used Python 3.0 for code development[9]

### 1.2.3 XGBoost

We would like to use the tensorflow decision forests to do the regression tree. However, Yggdrasil Decision Forest C++ libraries are not successful to install in researcher and client's machine. We cannot run tensorflow decision forests to forecast the signal.

Client is accepted to use xgboost instead of tensorflow decision forests to do the regression tree. XGBoost is able to gain the best annual return and lowest RMSE in training data.

## 1.3 Advancing knowledge and Skills

- Able to handle Tensorflow related project
  - Finished Tensorflow official's tutorial[4]
  - Finished Udacity: intro to tensorflow for deep learning course [Figure 2]
- Able to handle python development
  - Finished Datacamp's python course [Figure 3]

## 2 Project Management

### 2.1 Forecast and Actual Delivery

Figure 4 and [10] showing the forecast timeline and deliverable versus actual timeline and deliverable. The actual timeline followed the forecast and we delivered more deliverables.

For example, in the 1.0 tasks section, Timepoint 1 on 14/02/2021, we expect to do literature review and research for algorithms. However, we also learnt tensorflow and python at the same time. In the 2.0 tasks section, Timepoint 2 on 01/05/2021, we plan to develop and train 4 algorithms and we develop and train 5 algorithms. Timepoint 3 on project end, we complete most of the task except the Viva.

## 2.2 Project Management Method Techniques and tools

Clients suggest using simple and easy authentication tools for project communication. Both of us are google users so we decided to use it for project management. Researchers need to handle all tasks and tickets manually.

Google spreadsheet is free and easy to use if the user knows how to use excel. We use it as a Gantt chart and Issue ticketing system.

### Gantt chart(google spreadsheet)

Gantt chart is a bar chart that provides a visual view of tasks scheduled over time. It can show the start and end date of the project, also each task period and their dependencies.

Most of the advanced Gantt chart software includes functions like creating task dependencies, adding milestones, and identifying the critical path of a project. However, google spreadsheet does not include those functions and need to be created by the user.

Since the project only involves researchers and clients, a Google spreadsheet as Gantt chart in this project is easy to manage if the team and the scope of the project is small.

### Issue ticketing system(google spreadsheet)

Issue ticketing system is part of a helpdesk application. Ticketing management system helps users manage request and support cases. We use it as messaging apps for communication and comments.

If the project is more than 3 people, I would not recommend using google spreadsheet as gantt chart and issue ticketing system because it does not have task assigning functionality and dashboards are complicated to build.

### Github for code version control

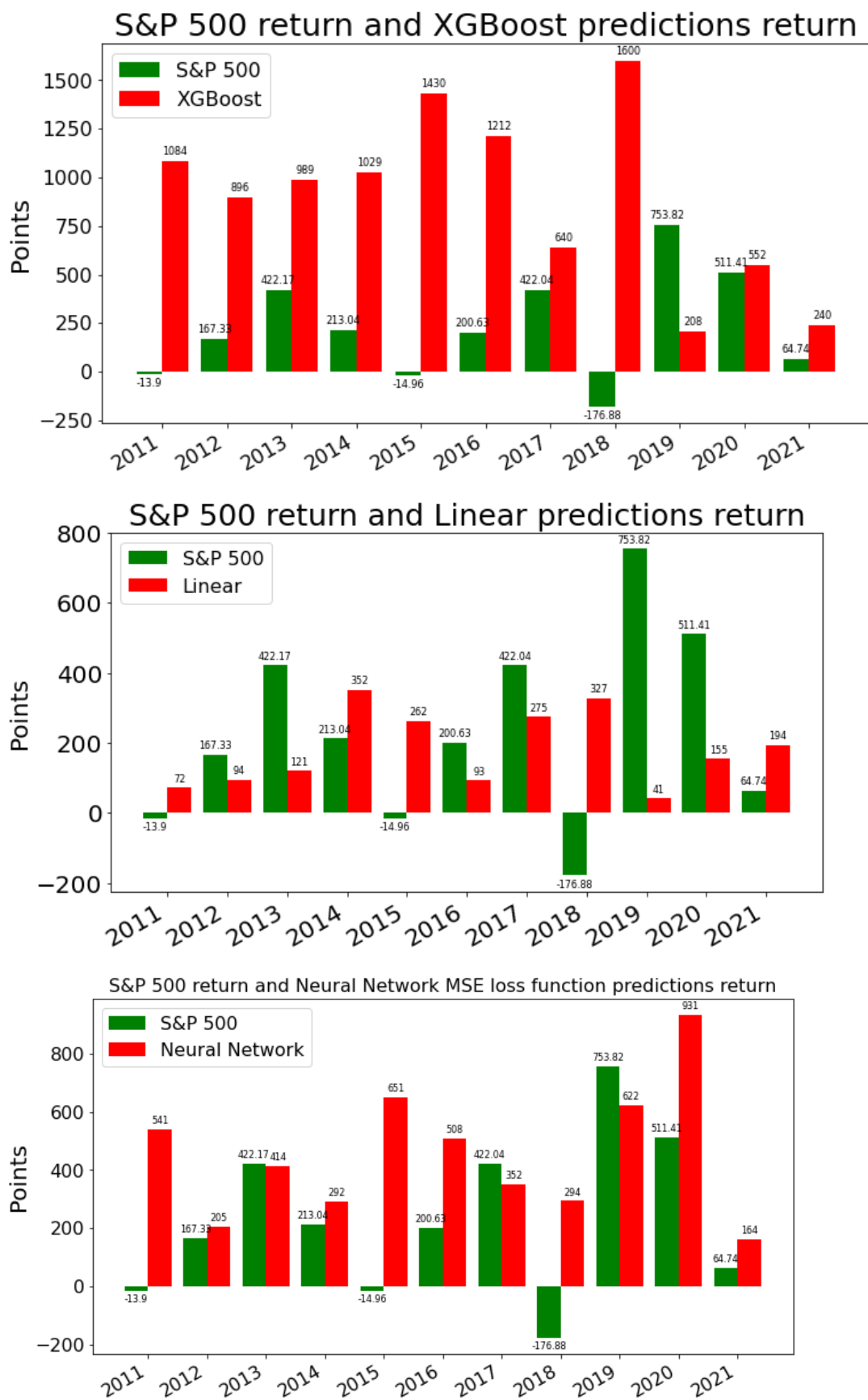
Git is a version control application that allows users to keep track of the work and help to easily explore the changes made. Such as data, graphics and coding scripts. The final coding is uploaded to github. [9]

As a programmer, I always recommend that all programming projects use github to do the version control and document because it is easy to maintain and contribute to open source projects.

## Reference

- [1] Omer Berat Sezer, Murat Ozbayoglu (2018) Algorithmic Financial Trading with Deep Convolutional Neural Networks: Time Series to Image Conversion Approach, Viewed 29 June 2021  
[https://www.researchgate.net/publication/324802031\\_Algorithmic\\_Financial\\_Trading\\_with\\_Deep\\_Convolutional\\_Neural\\_Networks\\_Time\\_Series\\_to\\_Image\\_Conversion\\_Approach](https://www.researchgate.net/publication/324802031_Algorithmic_Financial_Trading_with_Deep_Convolutional_Neural_Networks_Time_Series_to_Image_Conversion_Approach)
- [2] Haoming Li, Xhijun Yang and Tianlun Li (2014) Algorithmic trading strategy based on massive data mining, Viewed 19 June 2021  
<http://cs229.stanford.edu/proj2014/Haoming%20Li.%20Tianlun%20Li.%20Zhijun%20Yang.%20Algorithmic%20Trading%20Strategy%20Based%20On%20Massive%20Data%20Mining.pdf>
- [3] Fre Imbert(2018) US stocks post worst year in a decade as the S&P 500 falls more than 6 % in 2018, Viewed 19 June 2021  
<https://www.cnbc.com/2018/12/31/stock-market-wall-street-stocks-eye-us-china-trade-talks.html>
- [4] TensorFlow Core Tutorials (2021) available at: <https://www.tensorflow.org/tutorials>  
Accessed at:29 June 2021
- [5] Github (2021) Available at :  
[https://github.com/kinzia/PROM02\\_FYP/blob/main/01\\_linear/TF\\_linear.ipynb](https://github.com/kinzia/PROM02_FYP/blob/main/01_linear/TF_linear.ipynb) Accessed: 29 June 2021
- [6] Github (2021) Available at :  
[https://github.com/kinzia/PROM02\\_FYP/blob/main/03\\_NeuralNetwork\\_mse\\_loss\\_function/NeuralNetwork\\_model\\_1.ipynb](https://github.com/kinzia/PROM02_FYP/blob/main/03_NeuralNetwork_mse_loss_function/NeuralNetwork_model_1.ipynb) Accessed: 29 June 2021
- [7] Github (2021) Available at :  
[https://github.com/kinzia/PROM02\\_FYP/blob/main/04\\_NeuralNetwork\\_model\\_custom\\_loss\\_function/NeuralNetwork\\_custom\\_loss\\_function.ipynb](https://github.com/kinzia/PROM02_FYP/blob/main/04_NeuralNetwork_model_custom_loss_function/NeuralNetwork_custom_loss_function.ipynb) Accessed: 29 June 2021
- [8] Github (2021) Available at :  
[https://github.com/kinzia/PROM02\\_FYP/blob/main/05\\_NeuralNetwork\\_emsembling\\_model/NeuralNetwork\\_emsembling.ipynb](https://github.com/kinzia/PROM02_FYP/blob/main/05_NeuralNetwork_emsembling_model/NeuralNetwork_emsembling.ipynb) Accessed: 29 June 2021
- [9] Github (2021) Available at: [https://github.com/kinzia/PROM02\\_FYP](https://github.com/kinzia/PROM02_FYP) Accessed: 29 June 2021
- [10] Github (2021) Available at:  
[https://github.com/kinzia/PROM02\\_FYP/blob/main/PlanningReviewVsActualGanttChart%20-%20Automatic%20Gantt%20Chart.pdf](https://github.com/kinzia/PROM02_FYP/blob/main/PlanningReviewVsActualGanttChart%20-%20Automatic%20Gantt%20Chart.pdf) Accessed: 30 June 2021

Figure 1



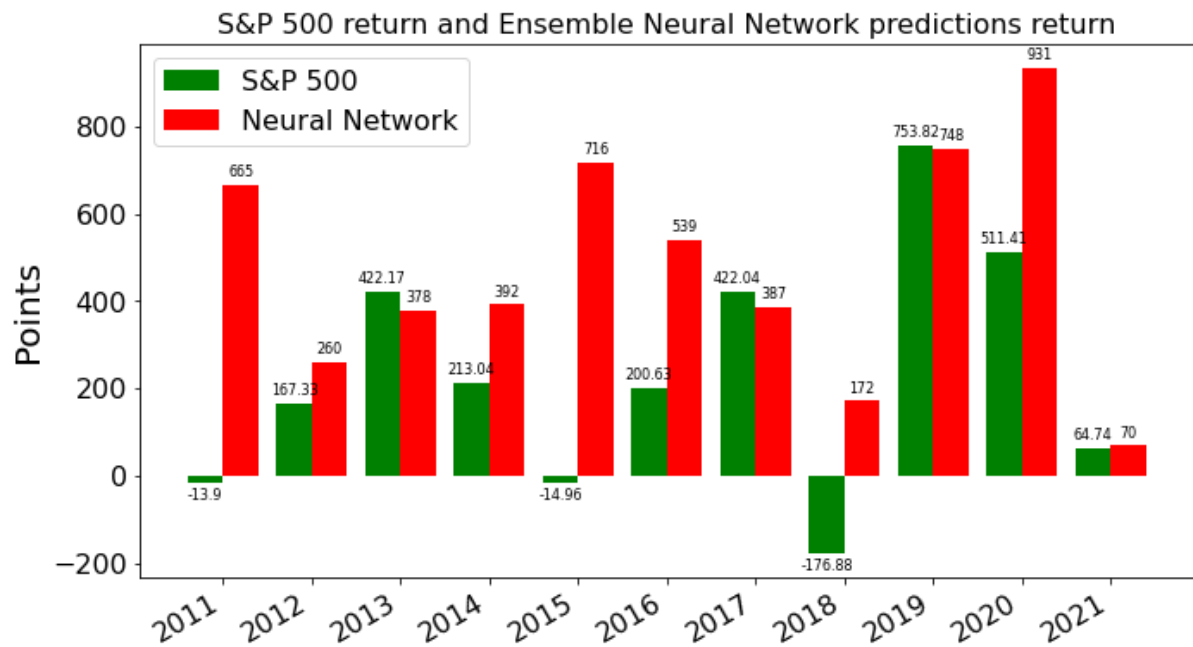
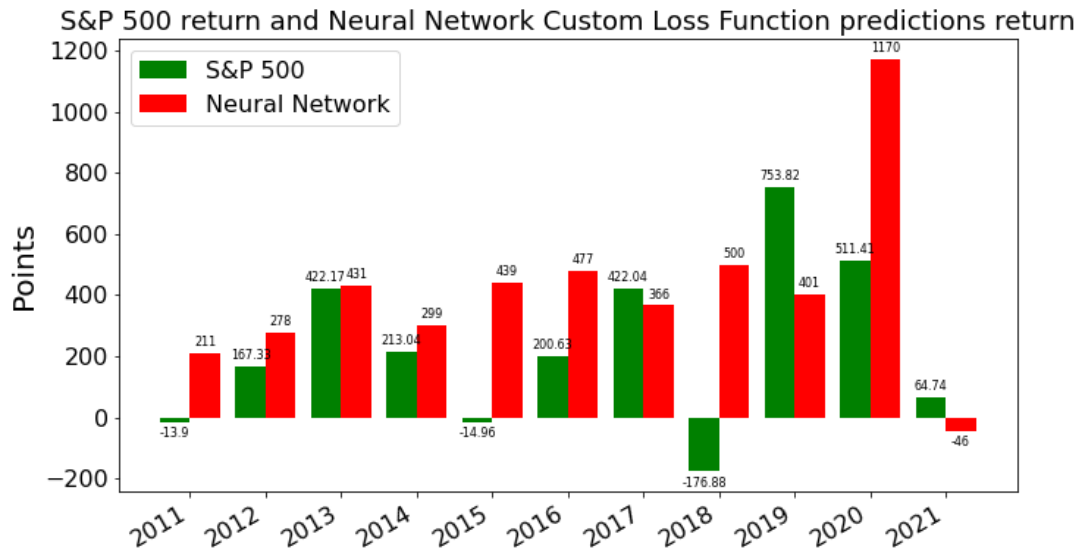






Figure 2

29/06/2021

Intro to TensorFlow for Deep Learning - Udacity

  
 HOME  
 HELP  
 CATALOG

# Intro to TensorFlow for Deep Learning

RESUME LESSON 11

Latest Activity: Lesson 11 - Introduction to TensorFlow Lite

100% VIEWED

LESSON 1

**Welcome to the Course**

Welcome to the course! Say hello to your instructors and get an overview of the program.

100% VIEWED

SHRINK CARD

Introduction to Machine Learning

COMPLETED

Your First Model - Fashion MNIST

COMPLETED

Introduction to CNNs

COMPLETED

Going Further With CNNs

COMPLETED

Transfer Learning

COMPLETED

Saving and Loading Models

COMPLETED

Time Series Forecasting

COMPLETED

NLP: Tokenization and Embeddings


COMPLETED

NLP: Recurrent Neural Networks

COMPLETED

Introduction to TensorFlow Lite

COMPLETED

 SETTINGS


 LOGOUT

Figure 3

29/06/2021

Introduction to Python | DataCamp

Learn

Search Catalog

...

INTERACTIVE COURSE

Introduction to Python

Practice Now

Replay Course

Bookmark

4 hours

11 Videos

57 Exercises

3,314,811 Participants

4,700 XP

Course Description

Python is a general-purpose programming language that is becoming ever more popular for data science. Companies worldwide are using Python to harvest insights from their data and gain a competitive edge. Unlike other Python tutorials, this course focuses on Python specifically for data science. In our Introduction to Python course, you'll learn about powerful ways to store and manipulate data, and helpful data science tools to begin conducting your own analyses. Start DataCamp's online Python curriculum now.

1 Python Basics FREE 100%

An introduction to the basic concepts of Python. Learn how to use Python interactively and by using a script. Create your first variables and acquaint yourself with Python's basic data types.

VIEW CHAPTER DETAILS

Completed

2 Python Lists 100%

Learn to store, access, and manipulate data in lists: the first step toward efficiently working with huge amounts of data.

VIEW CHAPTER DETAILS

Completed

3 Functions and Packages 100%

You'll learn how to use functions, methods, and packages to efficiently leverage the code that brilliant Python developers have written. The goal is to reduce the amount of code you need to solve challenging problems!

VIEW CHAPTER DETAILS

Completed

4 NumPy 100%

NumPy is a fundamental Python package to efficiently practice data science. Learn to work with powerful tools in the NumPy array, and get started with data exploration.

VIEW CHAPTER DETAILS

Completed

Statement of Accomplishment

View

This course is part of these tracks:

Data Scientist with Python

Python Fundamentals

Hugo Bowne-Anderson

Data Scientist at DataCamp

Hugo is a data scientist, educator, writer and podcaster formerly at DataCamp. His main interests are promoting data & AI literacy, helping to spread data skills through organizations and society and doing amateur stand up comedy in NYC. If you want to know what he likes to talk about, definitely check out [DataFramed](#), the DataCamp podcast, which he hosted and produced.

See More

COLLABORATOR(S)

Vincent Vankrunkelsven

Patrick Varilly

Filip Schouwenaars

<https://learn.datacamp.com/courses/intro-to-python-for-data-science>

1/1



Figure 4

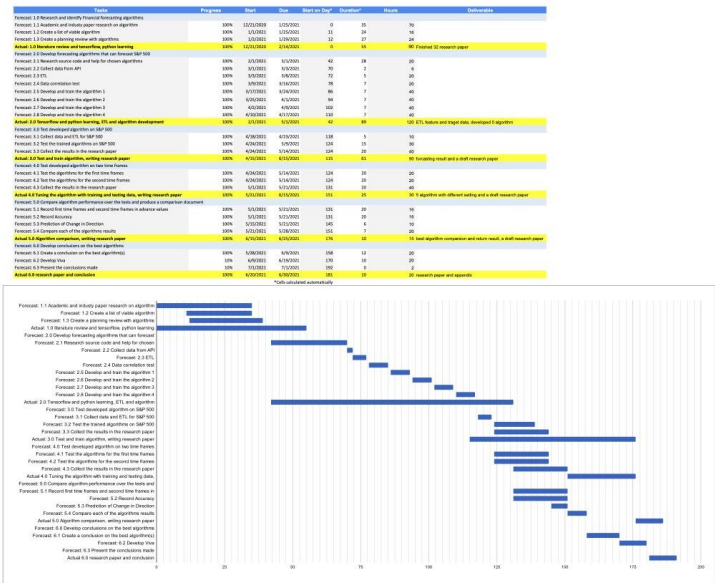


Figure 5

My Data science final project is finished! Thanks for your help!!

收件匣

✕

✕

✕

✕

Kevin Chow <kinzia@gmail.com>

下午1:51 (8 小時前)

☆

↶

⋮

寄給 Ted

Hi Ted,

Thanks for your support on the project and trading strategy. It drive the project on the right track and a good prediction result. Source code and the research paper is available at [https://github.com/kinzia/PROM02\\_FYP](https://github.com/kinzia/PROM02_FYP)

If you have any questions on this project please let me know.

Thanks!

Kevin

Ted Chan

下午2:35 (7 小時前)

☆

↶

⋮

寄給 我

Hi Kevin,

I am surprised that Tensorflow could generate a good annual return, and also satisfy the prediction result.

Would you please share the google spreadsheet gantt chart template, and ticketing system template that we use on this project? I might want to use it on another project as well.

Thanks.

Regards,

Ted

\*\*\*