



Automated news reading: Stock Price Prediction based on Financial News

指導教授：戴天時 組員：賴昱均、吳元魁、劉雅淳、蕭宛榆



01 研究動機

02 研究方法

03 研究結果

04 商業應用

05 總結

研究動機



新聞



金融科技





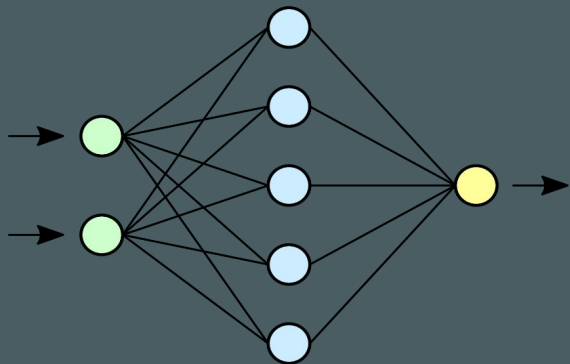
研究方法

架構

文字特徵



模型



預測股市



文字特徵

如何讓模型看懂文字?



N-gram

Uni-Gram

I

am

feeling

hungry

Bi-Gram

I am

am feeling

feeling hungry

Tri-Gram

I am feeling

am feeling hungry

I am feeling hungry

Text
feature

TF-IDF

TF*IDF

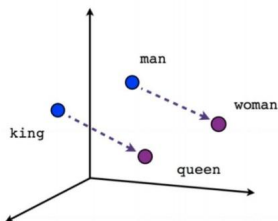
$TF \cdot IDF = \text{TERM FREQUENCY} \cdot \text{INVERSE DOCUMENT FREQUENCY}$

TERM FREQUENCY=
THE AMOUNT OF TIMES A
TERM APPEARS IN A
DOCUMENT

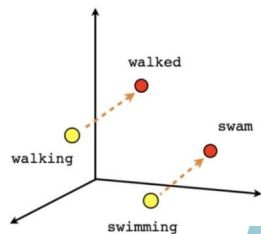


INVERSE DOCUMENT
FREQUENCY=
A MEASURE OF WHETHER A
TERM IS RARE OR COMMON
IN A COLLECTION OF
DOCUMENTS.

Word
Vector



Male-Female



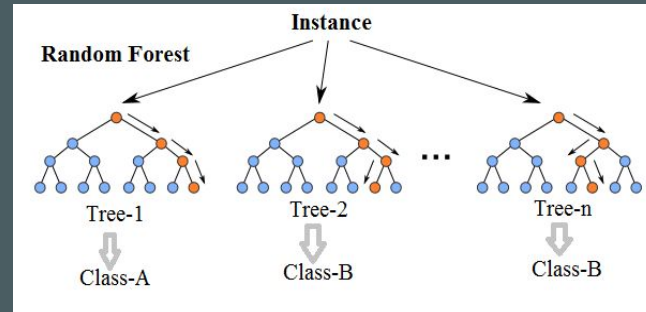
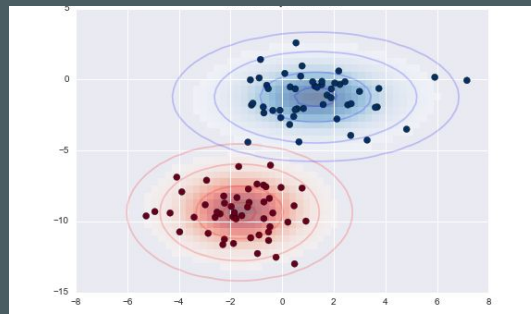
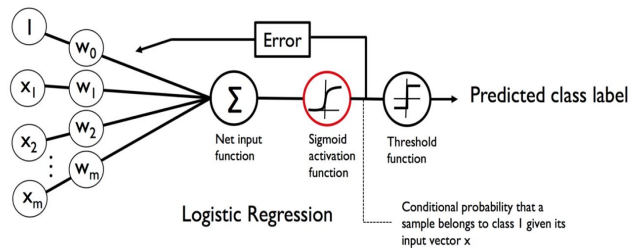
Verb tense



模型

Traditional ML Model

Deep Neural Network



Logistic Regression



Naive Bayes Classifier

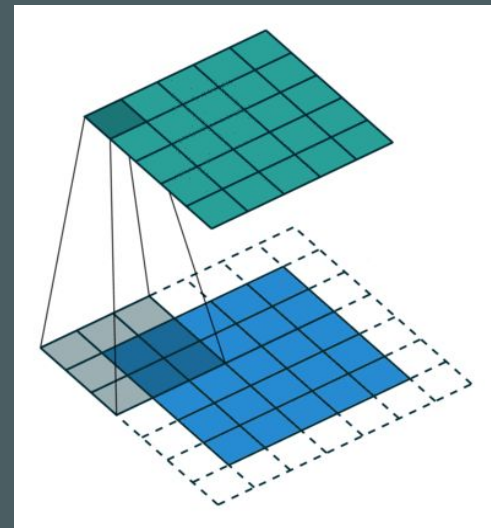
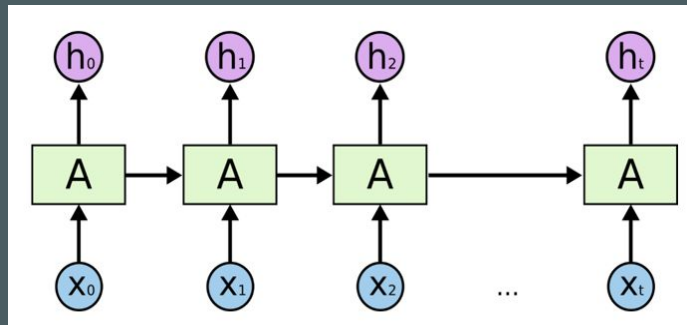
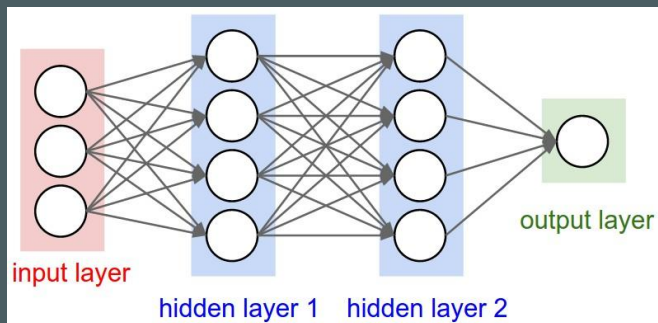


Random Forest

模型

Traditional ML Model

Deep Neural Network



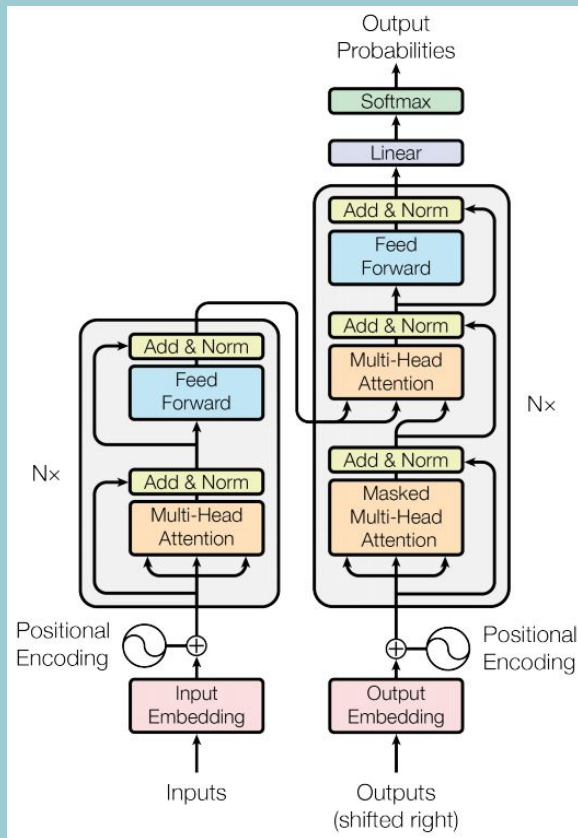
 Fully Connected Layer

 LSTM

 CNN



Transformer



BERT

1. Google 所提出的語言代表模型。
2. 在許多自然語言的任務中達到最佳的結果。
3. 將模型訓練在兩個任務上
 - a. 克漏字填空
 - b. 相鄰句子預測

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. **Bert: Pre-training of deep bidirectional transformers for language understanding**. arXiv preprint arXiv:1810.04805, 2018.

研究結果



大盤預測

道瓊指數 DJIA

	fc	cnn	lstm
wv	0.462	0.489	0.508
	logistic	nb	rf
1gram	0.429	0.484	0.495
1tfidf	0.508	0.513	0.5
2gram	0.56	0.487	0.511
2tfidf	0.577	0.542	0.566

Bert	
ep1	0.51
ep2	0.54
ep6	0.51
ep1_fr	0.51
ep2_fr	0.51
ep6_fr	0.51

SPY

	fc	cnn	lstm
wv	0.481	0.513	0.524
	logistic	nb	rf
1gram	0.502	0.508	0.524
1tfidf	0.524	0.521	0.529
2gram	0.55	0.534	0.532
2tfidf	0.521	0.519	0.497

Bert	
ep1	0.52
ep2	1
ep6	1
ep1_fr	0.52
ep2_fr	0.52
ep6_fr	0.52

個股預測

當日漲跌預測

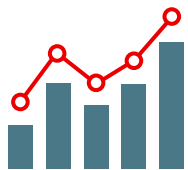
	fc	cnn	lstm
wv	0.524	0.517	0.515
	logistic	nb	rf
1gram	0.523	0.513	0.548
1tfidf	0.516	0.515	0.542
2gram	0.522	0.514	0.551
2tfidf	0.516	0.51	0.537

Bert
0.52

十天後漲幅預測

	fc	cnn	lstm
wv	0.546	0.536	0.556
	logistic	nb	rf
1gram	0.564	0.544	0.592
1tfidf	0.556	0.557	0.575
2gram	0.569	0.526	0.597
2tfidf	0.566	0.552	0.578

Bert
0.56



商業應用



於下單介面提醒



點選以了解更多預測原理...

今日上漲機率：57.7%

Line 官方帳號每日播送



< 永豐金證券

新功能上線囉！
✓每日美股大盤盤勢預測
✓每日美股個股走勢預測
採用機器深度學習分析而成，
準確率最高可達59.9%，
提供投資人另一項參考指標！
輸入“1”以查看今日預測
輸入“2”以了解模型原理

下午11:41

已讀 上午8:53 1

S&P500 今日上漲機率57.7%
AAPL 今日上漲機率58.5%
MSFT 今日上漲機率55.9%

更多個股漲跌預測，請輸入美股股票代碼
想了解更多模型分析原理，請輸入“2”

上午8:53

已讀 下午8:51 AMZN

AMZN 今日下跌機率57.3%

下午8:51

永豐金證券

台大課程 美股上漲預測表

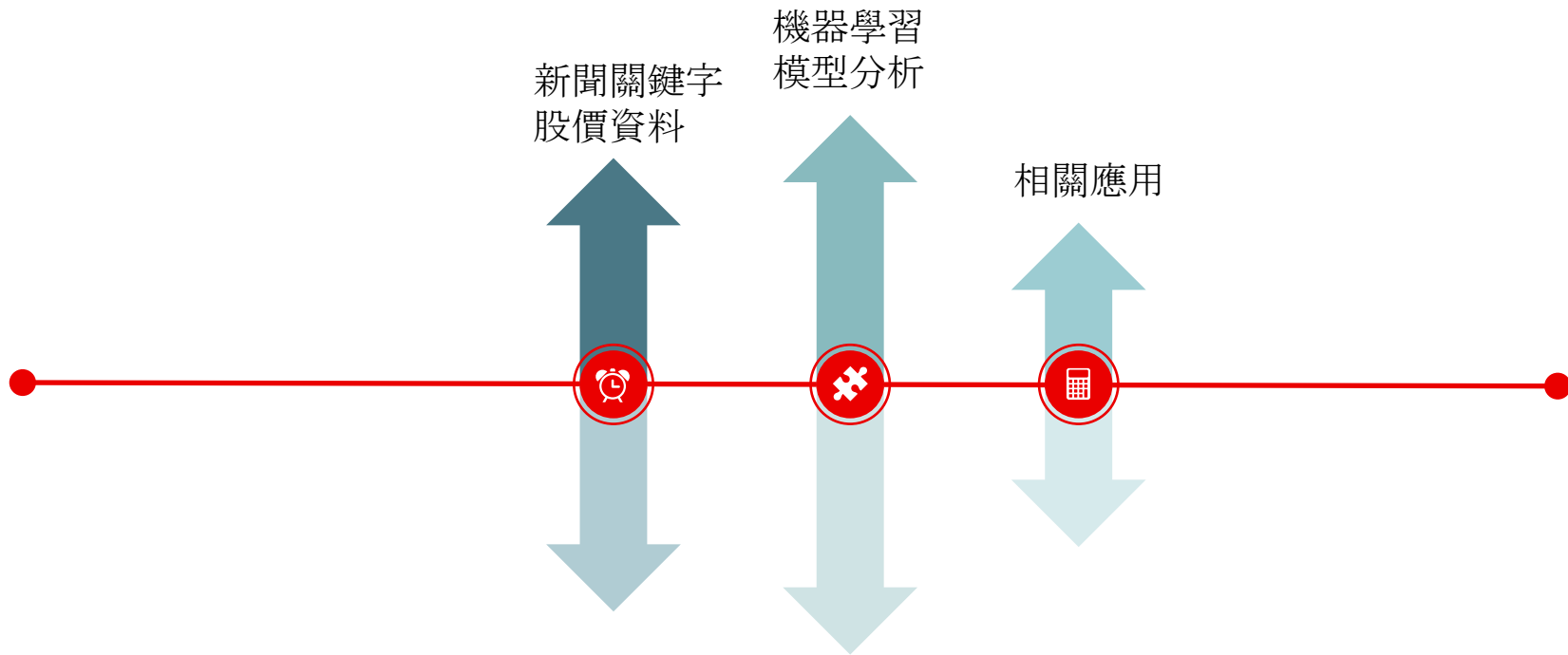
點我選擇最佳標的

Symbol	Price	Product	Rise Prob (1 day)	Rise Prob (10 days)
DJIA	33823.45	index	-20.82	70.93
NASDAQ	14161.35	index	38.34	-78.91
TSLA	615.56	stock	-63.8	47.04
AAPL	131.65	stock	-80.39	-75.1
NVDA	745.9	stock	-46.23	-35.65
AMZN	3487.73	stock	-41.76	3.83
SPY	421.97	ETF	33.33	42
ARKK	118.5	ETF	-53.95	19

預測日期：2021/06/21

資料僅供參考，以往績效不保證未來獲利

總結



A person in a dark suit is holding a tablet computer. The tablet screen shows a candlestick chart with green and red bars and a yellow trend line. The background is a dark, stylized city skyline at night, with a prominent skyscraper in the center. Overlaid on the right side of the image is a larger, semi-transparent candlestick chart with green and red bars and a yellow trend line. The text "Thank You" is written in a large, white, sans-serif font in the center-right of the image.

Thank You

分工表

1. 吳元魁: 資料收集、程式撰寫、實驗數據、投影片(研究方法、結果)
2. 賴昱均: 投影片(研究動機)、投影片設計
3. 劉雅淳: 投影片(商業應用、總結)、投影片設計
4. 蕭宛榆: 投影片設計、影片剪輯