

Stock data analysis



匯報
學生

105033467
譚詩毅



指導老師：邱銘傳教授

002785股票不錯
000593股票不錯
603505股票不錯
002806股票不錯
300332股票不錯
000507股票不錯
300334股票不錯
002611股票不錯
603903股票不錯
000636股票不錯



CONTENTS

01

Introduction

02

Big data

03

Data mining

04

Demo

05

Conclusion & future

01

Introduction

股票分析

股票分析真的是個很大問題，研究的人很多，講故事的騙子也很多。那到底什麼是股票分析，股票分析要做什麼，真的是百家爭鳴，百花齊放。

但是既然是交易，就一定會有滿足經濟學規律的地方，逃不了買賣二字。但是很不巧，我的工程經濟也是這學期才修的，但是搞不好就是因為沒什麼基礎，所以大家應該都能聽懂。

introduction—where & when

何時何地

如果僅僅是考慮股票數據分析，那就是股票交易的時候使用。但是我覺得一個方法和背後的思想應該是值得廣泛推廣的。就例如我們所學的BPR 方法，生成會有瓶頸，人生有沒有瓶頸。同樣的，在解決這個題目的過程中，一些收穫和處理思路也可以運用到其他問題的解決上。

introduction—who

何人使用

想要一夜暴富么

想要躺著賺錢么

想要人生勝利么

歡迎加入安利



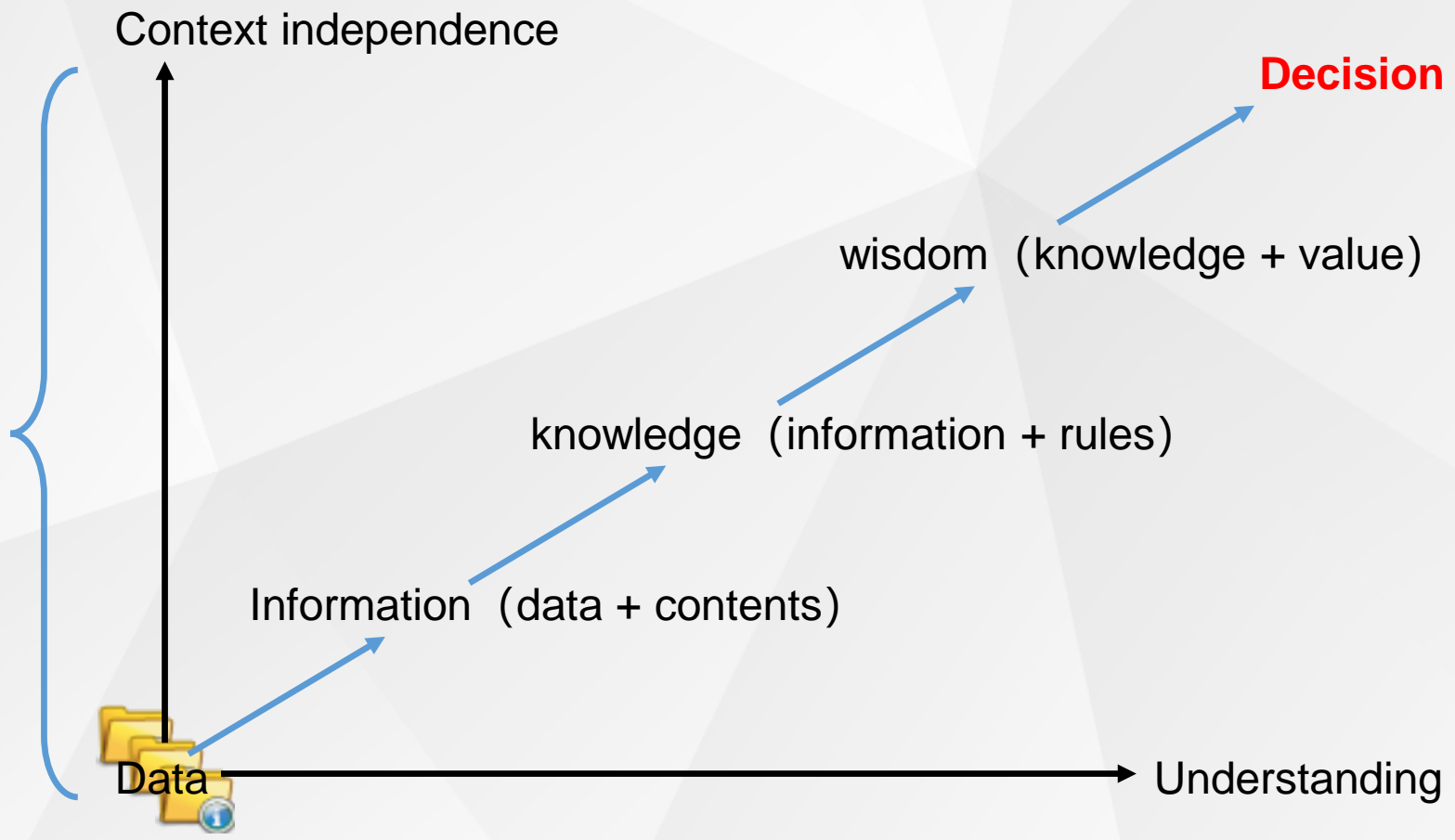
使用股票數據分析



introduction—why

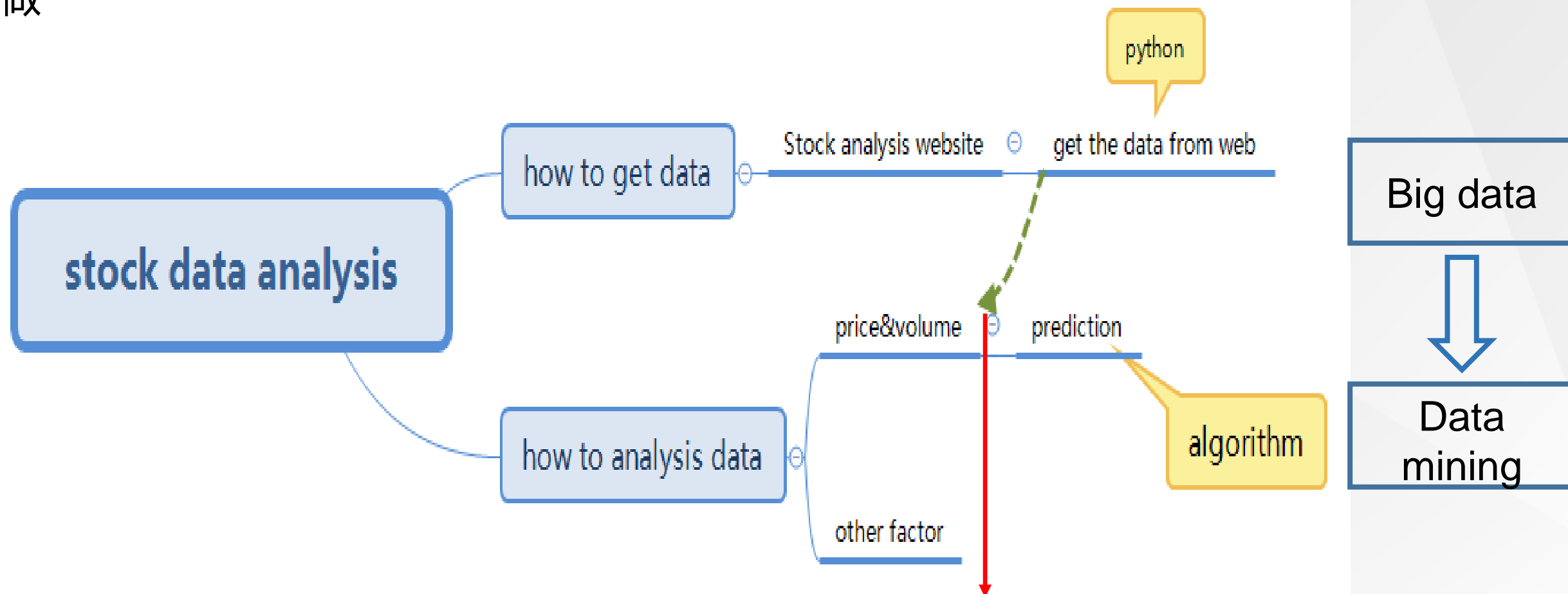
為何 DIKWD model (data→information → knowledge → wisdom → decision)

內線炒股
靈魂炒股 
拍腦袋炒股
科學炒股



introduction—how

怎麼做



Breakdown要更細緻，一個簡單的綠線，數據溝通就寫了一天的code，最開始用了50多行完成

02

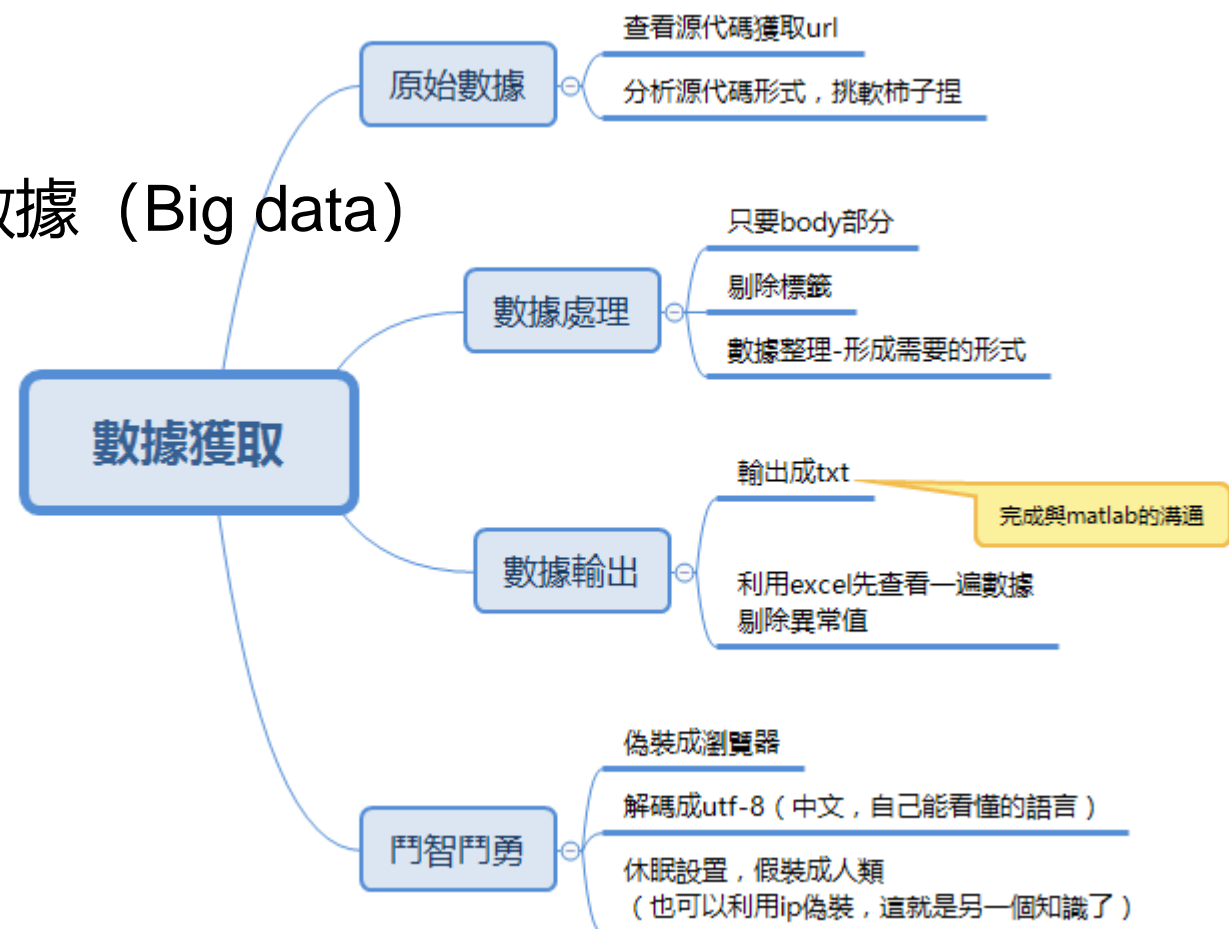
Big data

Big data

什麼是Big data? 3V

要進行數據分析，首先要有數據，大量的數據 (Big data)

- 如何獲取數據
- 如何自動獲取數據
- 分析數據特征，是否是自己需要的數據
- 如何篩選數據
- 如何存儲數據
- 如何利用數據



Big data-網頁分析

"http://quote.stockstar.com/stock/ranklist_a_3_1_3.html"

首页 股票 行情 数据 研报 基金 债券 期指 期货 黄金 外汇 银行 视频

我的证券之星 [登录] [注册] | 设为首页 | 收藏本站



有深度的财经资讯 有重磅的财经头条



行情中心 QUOTE.STOCKSTAR.COM 最新、最全、最精

代码/简称/拼音 查询

- 沪深市场
- 市场分类
 - 沪市A股
 - 深市A股
 - 沪市B股
 - 深市B股
 - 中小板
 - 创业板
 - A股市场
 - B股市场
 - A+B股
 - 新股
 - 沪港通(沪股通)
 - 沪市风险警示板
 - 深市退市整理板
 - 三板
- 标准行业
- 概念板块
- 地域板块
- 指数列表
- 综合排行
 - 全球
 - 基金
 - 期货
 - 债券
 - 外汇
 - 黄金
 - 自选股

当前位置: 首页 - 行情 - 沪深市场 - 沪深A股

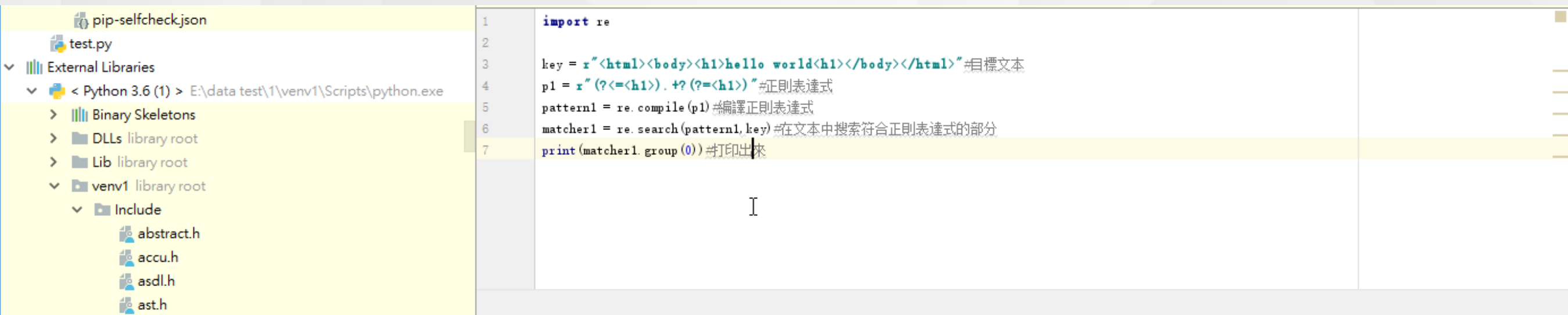
返回行情中心

沪深A股												
代码	简称	最新价	涨跌幅	涨跌额	5分钟涨幅	成交量(手)	成交额(万元)	换手率	振幅	量比↑	委比	市盈率
000818	方大化工	12.98	4.26%	0.53	0.70%	165248.98	21110.48	2.39%	4.58%	0.01	0.48	76.73
000975	银泰资源	13.36	-0.67%	-0.09	0.30%	41418.58	5523.22	0.47%	2.45%	0.01	-0.14	191.18
000822	山东海化	9.05	-0.55%	-0.05	0.11%	186119.10	16851.20	2.08%	2.09%	0.02	0.42	7.35
000825	太钢不锈	4.96	0.40%	0.02	0.00%	334294.19	16577.14	0.59%	1.42%	0.02	-0.48	21.95
000830	鲁西化工	15.92	4.67%	0.71	0.00%	479206.82	75706.05	3.27%	5.72%	0.02	-0.05	24.27
002504	*ST弘高	6.27	-5.00%	-0.33	0.00%	6379.00	399.96	0.15%	0.00%	0.03	-1.00	70.61
002305	南国置业	4.62	-0.22%	-0.01	-0.22%	52496.00	2423.70	0.32%	1.08%	0.04	-0.57	394.20
000839	中信国安	9.59	3.34%	0.31	-0.10%	1383727.30	132994.56	3.53%	5.60%	0.05	0.37	172.23
000889	茂业通信	13.34	-0.45%	-0.06	-0.22%	5410.50	719.56	0.09%	2.54%	0.05	0.38	45.39
000600	建投能源	7.71	0.39%	0.03	-0.13%	71200.57	5481.72	0.65%	1.82%	0.06	0.02	15.33
002421	达实智能	6.05	1.51%	0.09	0.67%	82919.11	4970.26	0.55%	1.51%	0.06	-0.32	72.33
002438	江苏神通	7.08	0.28%	0.02	0.28%	5514.02	389.91	0.20%	1.13%	0.06	0.15	57.26
300686	智动力	22.70	-0.26%	-0.06	0.40%	9389.30	2131.44	3.00%	2.02%	0.06	0.69	—
002135	东南网架	7.20	1.12%	0.08	-0.14%	32556.00	2334.58	0.49%	1.83%	0.07	-0.20	47.43
002165	红宝丽	6.09	0.00%	0.00	0.16%	12371.38	750.16	0.23%	1.31%	0.07	-0.88	48.20
002549	凯美特气	6.54	2.51%	0.16	0.62%	24628.49	1594.33	0.40%	2.98%	0.07	-0.15	176.76
300041	回天新材	10.02	0.50%	0.05	0.10%	17427.00	1743.09	0.67%	0.80%	0.07	-0.46	21.72
300311	任子行	14.00	0.65%	0.09	0.00%	14295.03	1999.52	0.60%	1.22%	0.07	-0.40	147.74
002729	好利来	61.96	0.02%	0.01	-0.03%	1246.00	771.08	0.19%	0.90%	0.08	0.90	266.24
300682	朗新科技	18.31	-0.33%	-0.06	-0.11%	14345.66	2644.42	3.19%	2.99%	0.08	-0.63	—
000707	双环科技	5.54	0.91%	0.05	0.18%	15803.67	870.92	0.34%	1.28%	0.09	-0.26	19.99
000905	厦门港务	10.54	-0.19%	-0.02	-0.28%	23398.25	2465.35	0.44%	1.04%	0.09	-0.26	65.65
000906	浙商中拓	7.29	1.11%	0.08	0.14%	15045.08	1091.47	0.35%	1.25%	0.09	0.01	20.61
002104	恒宝股份	8.77	0.57%	0.05	-0.11%	46604.86	4075.22	0.77%	1.26%	0.09	-0.42	37.80
002282	博深工具	14.09	0.21%	0.03	0.14%	15501.06	2183.79	0.81%	0.85%	0.09	-0.01	287.55
002395	双象股份	12.08	0.42%	0.05	0.17%	5075.00	610.28	0.28%	1.08%	0.09	0.00	138.34

Name	Status	Type	Initiator	Size	Time	Timeline - Start Time
ranklist_a_10_0_11...	200	doc...	Other	10.8...	181...	
common.css	200	styl...	ranklist a 10 0 11.html:10	(fro...	0 ms	
quotecommon.mi...	200	script	ranklist a 10 0 11.html:12	(fro...	0 ms	
sortpagetype.min.js	200	script	ranklist a 10 0 11.html:14	(fro...	0 ms	
StockSuggest.min....	200	script	ranklist a 10 0 11.html:59	(fro...	0 ms	
SS_dcs_tag.js	200	script	ranklist a 10 0 11.html:703	(fro...	0 ms	
s?z=stockstar&c=4...	200	script	ranklist a 10 0 11.html:707	372 B	280...	
s?z=stockstar&c=4...	200	script	ranklist a 10 0 11.html:708	372 B	542...	
bg_submenu2.png	200	png	ranklist a 10 0 11.html:593	(fro...	0 ms	
icon.gif	200	gif	ranklist a 10 0 11.html:593	(fro...	0 ms	
bg.gif	200	gif	ranklist a 10 0 11.html:593	(fro...	0 ms	
bg_title.png	200	png	ranklist a 10 0 11.html:593	(fro...	0 ms	
h2_title.png	200	png	ranklist a 10 0 11.html:593	(fro...	0 ms	
formelement.png	200	png	ranklist a 10 0 11.html:593	(fro...	0 ms	
bg_submenu.gif	200	gif	ranklist a 10 0 11.html:593	(fro...	0 ms	
setshareid.jsp	200	doc...	SS_dcs_tag.js:5	964 B	58 ms	
dcs.gif?d&dcsdat=1...	200	gif	Other	410 B	115...	
icon3.png	200	png	ranklist a 10 0 11.html:707	(fro...	0 ms	
h.js?d2eddf9b155b...	307		ranklist a 10 0 11.html:724	0 B	0 ms	
h.js?d2eddf9b155b...	304	script	http://hm.baidu.com/h.js?d2...	208 B	146...	
hm.gif?cc=0&ck=...	307		Other	0 B	53 ms	
hm.gif?cc=0&ck=...	307		Other	0 B	53 ms	
s?z=stockstar&c=1...	200	doc...	quotecommon.min.js:1	760 B	294...	
s?z=stockstar&c=5...	200	doc...	quotecommon.min.js:1	746 B	279...	
rank.ashx?type=a...	200	xhr	quotecommon.min.js:17	2.7 KB	1.17 s	
bg_fenye.png	200	png	quotecommon.min.js:21	(fro...	0 ms	
hm.gif?cc=0&ck=...	200	gif	http://hm.baidu.com/hm.gif...	299 B	170...	
hm.gif?cc=0&ck=...	200	gif	http://hm.baidu.com/hm.gif...	299 B	153...	
800-6020171116.jpg	200	jpeg	s?z=stockstar&c=51&top=1:1	(fro...	0 ms	
guanggao.png	200	png	s?z=stockstar&c=51&top=1:1	(fro...	0 ms	
hqpdzan120171124...	200	jpeg	s?z=stockstar&c=193&top=1:1	(fro...	0 ms	
guanggao.png	200	png	s?z=stockstar&c=193&top=1:1	(fro...	0 ms	

37 requests | 18.8KB transferred | Finish: 2.24 s | DOMContentLoaded: 974 ms | Load: 1.25 s

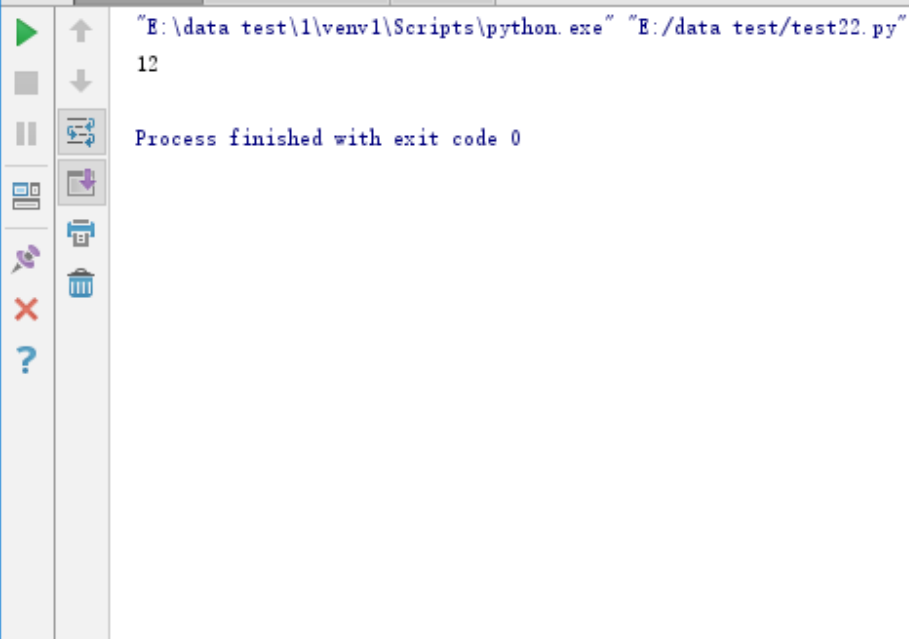
Big data-數據獲取-正則判斷



The screenshot shows an IDE with a file explorer on the left and a code editor on the right. The file explorer shows a project structure with files like `test.py` and a virtual environment `venv1`. The code editor contains the following Python code:

```
1 import re
2
3 key = r"<html><body><h1>hello world</h1></body></html>" #目標文本
4 p1 = r"(?<=<h1>). +?(?=<h1>)" #正則表達式
5 pattern1 = re.compile(p1) #編譯正則表達式
6 matcher1 = re.search(pattern1, key) #在文本中搜索符合正則表達式的部分
7 print(matcher1.group(0)) #打印出來
```

Run: test22 hello world test



The run console shows the execution of the Python script. The command executed is `"E:\data test\1\venv1\Scripts\python.exe" "E:/data test/test22.py"`. The output is `12`. The console also shows `Process finished with exit code 0`.

Big data-數據獲取-正則判斷

```
response = urllib.request.urlopen(request)
content = response.read().decode('gbk') # 讀取網頁內容，解碼gbk碼
print('get page successfully', page) # 顯示已經成功讀取的頁面編碼
pattern = re.compile('<tbody[\s\S]*</tbody>') # 利用正則表達式，讀取標籤<tbody class="tbody_right" id="datalist"></tbody>中內容
body = re.findall(pattern, str(content)) # re.findall返回符合要求的元素列表
pattern = re.compile('>(.*?)<') # 讀取標籤>和<中內容，即每隻股票內容
stock_page = re.findall(pattern, body[0]) # 從body中re.findall返回符合要求的元素列表
```

"E:\data test\1\venv\Scripts\python.exe" "E:/data test/test22.py"

```
[',', '000612', '焦作電力', '9.62', '6.42%', '0.58', '0.10%', '67806.72', '15770.89', '1.68%', '6.86%', '0.92', '-0.42',
'91.76', '600183', '生益科技', '17.76', '6.35%', '1.06', '0.23%', '417433.95', '72980.07', '2.87%', '7.90%',
'1.23', '-0.34', '22.71', '002842', '翔鷗鈷業', '43.87', '6.33%', '2.61', '0.05%', '37594.21', '16134.37',
'15.04%', '10.47%', '2.48', '-0.77', '67.48', '600549', '廈門鈷業', '26.25', '6.19%', '1.53', '0.00%', '541593
.85', '138807.67', '5.01%', '6.88%', '2.29', '-0.68', '54.10', '002837', '英維克', '20.99', '6.12%', '1.21',
'0.00%', '68582.00', '14209.23', '13.93%', '11.17%', '1.20', '-0.62', '77.78', '300497', '富祥股份', '35.81', '6.01%',
'2.03', '0.11%', '34508.34', '12216.32', '5.85%', '6.54%', '2.40', '-0.49', '19.82', '603599', '廣信股份', '19.30',
'5.99%', '1.09', '0.16%', '45194.50', '8562.54', '3.31%', '6.75%', '1.39', '0.28', '25.29', '603725', '天安新材',
'21.13', '5.97%', '1.19', '-0.24%', '72755.85', '15230.37', '19.84%', '12.04%', '2.35', '0.19', '—', '300729',
```

Big data-數據輸出

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	601313	江南嘉捷	55.63	57.57%	5.06	0.00%	178632.55	97270.88	4.50%	34.36%	14.38	1.00	336.74				
2	601326	秦港股份	6.57	16.24%	0.38	0.46%	567097.69	36666.05	10.16%	23.93%	0.00	-0.46	28.82				
3	603283	赛腾股份	17.59	16.10%	1.60	0.00%	548.71	96.52	0.00%	0.00%	46.61	1.00	0				
4	603799	华友钴业	88.31	10.94%	7.34	0.89%	216586.13	186257.54	7.98%	10.40%	2.32	-0.28	51.99				
5	600985	雷鸣科化	14.70	10.84%	1.34	0.00%	127308.38	18332.68	4.84%	10.92%	19.51	1.00	81.79				
6	600345	长江通信	29.88	10.61%	2.07	0.27%	81980.04	23639.61	4.14%	12.61%	7.90	-0.19	22.80				
7	000409	山东地矿	5.89	10.09%	0.54	0.00%	65174.95	3749.09	1.65%	8.79%	0.75	1.00	0				
8	000600	建投能源	8.65	10.05%	0.79	0.00%	395126.77	33040.45	3.62%	9.92%	0.65	1.00	17.20				

就绪

finaltest

计数: 38699

100%

03

Data mining

Data mining

什麼是Data mining? model

獲得數據后如何進行數據分析，關鍵就是構建模型

- 決策變量 (主要原因、次要原因)
- 目標函數 (多目標)
- 約束條件
- 更新方式
- 終止條件



Data mining-model

Model

股票篩選2

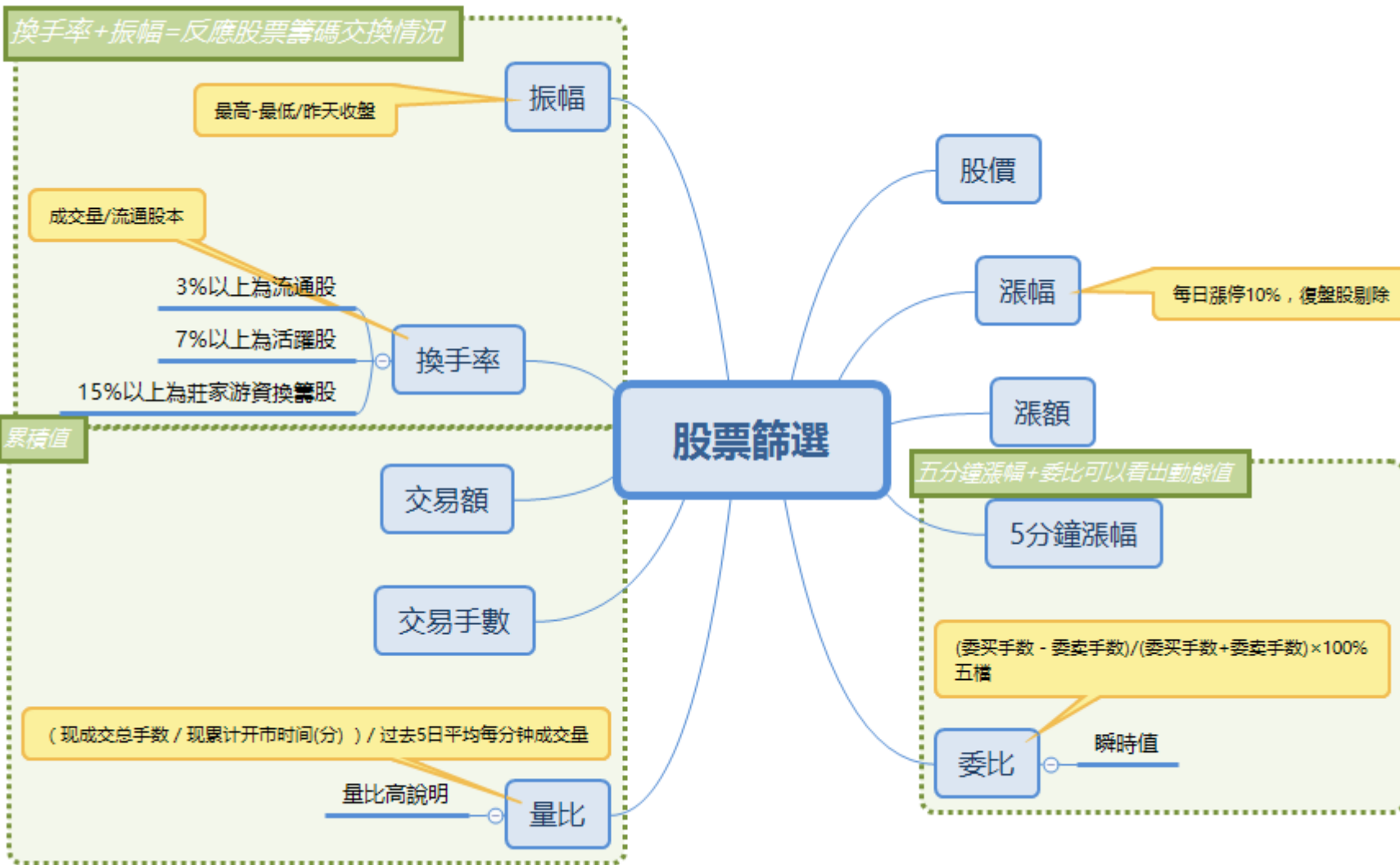
可能的價格± 10%

每個價格的可能性

迭代出有可能的價格波段

買賣點

有時候不是你能打出什麼牌，
而是怎麼把手上的牌打好



Data mining-決策、目標、限制→迭代

object: $z = w_1 * rand_1 * \text{stock increase per 5min} +$

$w_2 * rand_2 * \text{turnover rate} +$

$w_3 * rand_3 * [\text{stroke count} / (\text{volume of trade} * \text{stock price})]$

$rand_4 * \text{stock amplitude} +$

$rand_5 * \text{price-earnings ratio} / \text{stock increase}$

$\text{turnover rate} > 0.3$

$\text{stock increase} \in (-0.1, 0.1)$

```
編輯器 - C:\Users\T\Documents\MATLAB\dataatest2.m
+1 fun.m x Untitled14.m x test1.m x testfor20171217A.m x dataatest2.m x Untitled4.m x dataatest3.m x
78
79 %目标函数暂定z=20*rand1*(5分钟)+10*rand2(换手率)+rand3*((成交额/(成交手*股價))*(5分钟)*(换手率)
80 %%%%%%%%%%%%%%% -rand4*(振幅)+rand5*(市盈率/張幅)*(5分钟)
81 A=z(:,1);
82 B=unique(A);
83 fin=[];
84 finally=[];
85 for i=1:81
86 store=zeros(size(B));
87 tempa=[];
88 for i=1:length(B)
89 f=find(A==B(i));%找出同一隻股的參數
90 [r]=size(f);
91 num=r*1;
92 X=z(f,7);%成交額
93 Y=z(f,6)+0.0001;%成交量
94 tempa=X./Y;
95 tempb=z(f,2);%股價X
96 Tempa=tempa./tempb;
97 X=z(f,5);%5分鐘張幅
98 Y=z(f,8)+0.0001;%換手率
99 Tempb=X.*Y;
100 Tempc=Tempa.*Tempb;
101 X=z(f,12);%市盈率
102 Y=z(f,3)+0.1;%張幅
103 tempc=X./Y;
104 tempd=z(f,5);
105 Tempd=tempc.*tempd;
106 object=20*rand().*z(f,5)+10*rand().*z(f,8)+5*rand().*Tempc-rand().*z(f,3)+rand().*Tempd;
107 o=sum(object(1,:));
108 ob=o/num;
109 store(i,1)=z(f(1,1),1);
110 store(i,2)=ob;
111 %換手率
112 ch=sum(z(f,9));
113 change=ch/num;
114 store(i,3)=change;
115 end
```

04

Demo

Demo

002785股票不錯
000593股票不錯
603505股票不錯
002806股票不錯
300332股票不錯
000507股票不錯
300334股票不錯
002611股票不錯
603903股票不錯
000636股票不錯



05

Conclusion & future

Conclusion & future

Big data

——爬數據是一個鬥智鬥勇的過程

- 1、破解 (URL、decode)
- 2、偽裝 (瀏覽器、休眠)
- 3、自動化 (while, 跳出錯誤)
- 4、智能化 (正則表達、數據整理)
- 5、數據傳遞



Data mining

- 1、數據篩選
- 2、數學模型：
 - 決策變量 (價量, 經濟學)
 - 目標函數 (股池+個股波動分析)
 - 約束條件 (約束條件其實越少越好)
 - 其他因素 (中國大媽、黑天鵝)
- 3、算法, 類神經? (為什麼一定要模型)

Future

- 1、區塊鏈?
記賬, 被大家認可的結果 (粒子群), 優先輸出?
- 2、量化 (其他干擾因素)
- 3、心理學、經濟學

感謝聆聽

