Railway

——the report of Introduction to business computing

Team project : Real-time bond pricing-railway

Team members : 吉迎 吴琼琳 王蕊 盛竞逸

CONTENT

1 Description

② Background

Huge Gap

the amount of reference information available to those trading equities

those trading corporate bonds

Delay

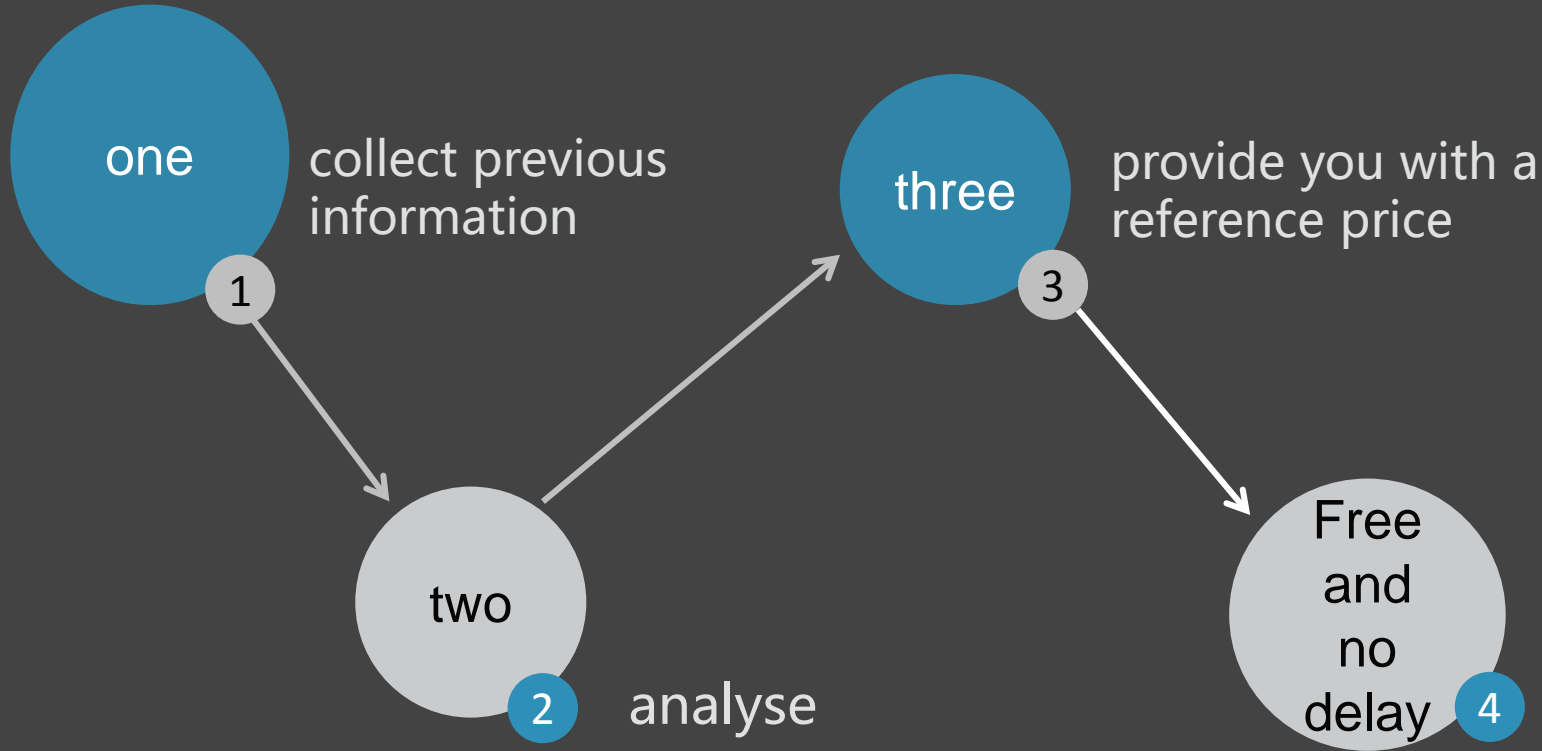Free access is available online with a 15 minute delay

Cost

more information need pay

Hinder

Accurate bond pricing is also hindered by lack of liquidity.

So our project was born

3 The process of programming

Before start:

>>we find there are countless missing values in the data

4	123.7866	94144	122.633	3650000		4	123.7866	963857	126.379	370000		2	126.4595	1113341
4	123.7866	105610	122.696	3445000		4	123.7866	106115	122.633	3650000		4	123.7866	975828
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	101.1022	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	101.1032	190064	101.07	25000		2	101.1022	NaN	NaN	NaN	NaN	NaN	NaN
2	101.1019	514299	101.086	500000		2	101.1032	520165	101.07	25000		2	101.1022	NaN
2	101.0774	514338	101.033	25000		2	101.1019	523476	101.086	500000		2	101.1032	529342
2	101.063	440599	100.9777	25000		2	101.0774	515487	101.033	25000		2	101.1019	524625

762675	915157	0.554384	5.75	10.00577	0	2.363	50000	3	101.4809
762676	915157	0.261769	5.75	10.00575	0	154.395	500000	2	101.4877
762677	915157	0.321389	5.75	10.00573	0	48.938	3700000	4	101.4464
762678	915157	0.351248	5.75	10.00571	0	9.511	2000000	4	101.4294
762679	915157	0.153716	5.75	10.00571	0	2.325	3700000	4	101.3553
762680									
762681									
762682									
762683									
762684									

ata and

Find the best algorithm in training the machine and predicting the price of bonds:
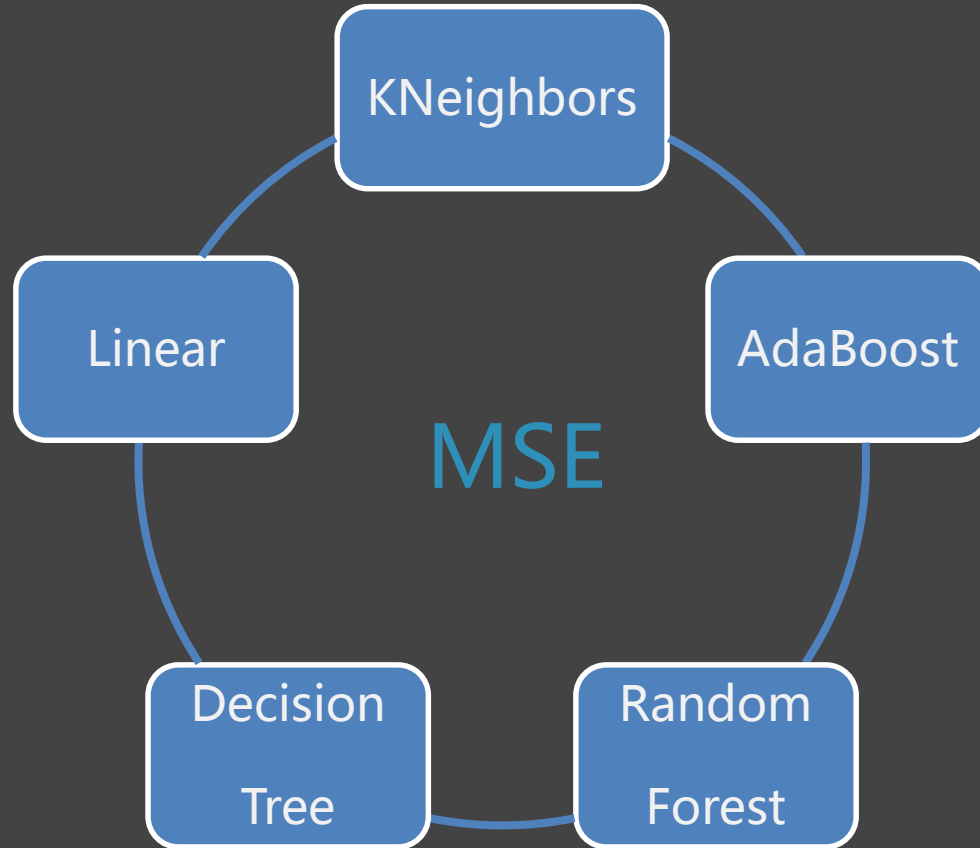
Detailed Procedures

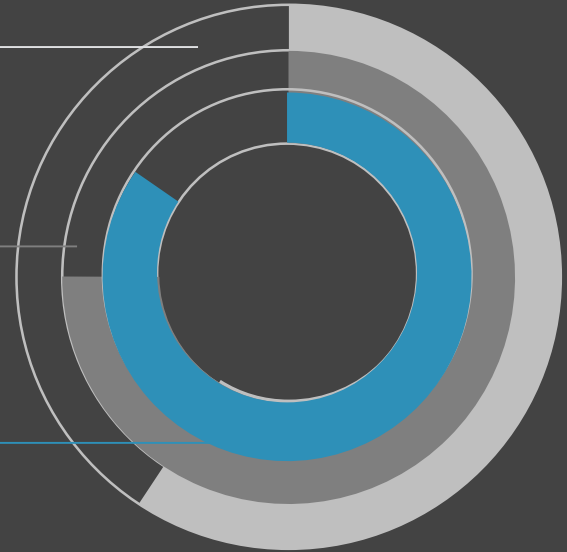Use panda to read the price data and prepare feature metric

First

Use holdout validation and train the model in each algorithm

Second

Evaluate the regression model and choose the best algorithm

Third

Evaluate the regression model

	KNeighbors Regressor	Linear Regressor	Decision Tree Regressor	Random Forest Regressor	AdaBoost Regressor
MSE	80.762049	0.930923	1.372639	0.655586	10.127853

Find the best algorithm to predict the price of the bond

ALL is done!

It's time to use the program and predict the bond price.

Here

We have a new data which has all detailed information but no trade price.

id	weight	current_c	time_to_m	is_callab	reporting	trade_siz	trade_typ	curve_bas
762679	0.733639	1.30278	1.090694	0	68.135	2250000	2	98.5218
762680	0.40357	8.5	7.472191	0	2.421	3000	2	119.2924
762681	0.41694	8.625	8.37855	1	25.469	1000000	4	103.9731
762682	0.012991	7	1.71599	0	12.188	25000	4	106.6809
762683	14.73066	5.625	1.241897	0	84.129	1000000	4	103.369
762684	0.755066	8	6.129066	1	437.017	1000001	2	102.5302
762685	0.022502	6	6.326842	0	3.347	10000	3	99.65152

Well
It finally runs well.

1	bond_id	trade_price	
2	762679	98.0921	
3	762680	116.183	
4	762681	104.386	
5	762682	106.71	
6	762684	102.076	
7	762685	100.402	
8	762686	105.45	
9	762687	103.079	
10	762688	106.135	
11	762689	98.1274	
12	762690	101.234	
13	762691	92.5543	
14	762692	100.94	
15	762693	104.273	
16	762694	110.418	
17	762695	99.8172	
18	762696	107.83	
19	762697	104.88	
20	762698	107.298	
21	762699	104.559	

Bond ID → Trade Price

predict.csv   Detailed information

result.csv   Exact price

4 Problems in programming

1 The missing value

2 Too huge file to operate

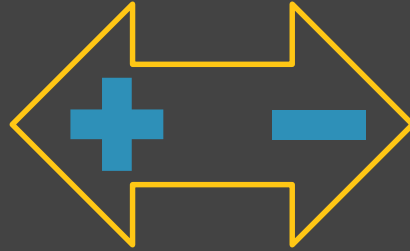**3** Application of the trained model

5 Reflection & Conclusion

Advantage

approximate real-time price fifteen minutes earlier

spare the money for buying the information

it's possible to predict the price because the bond market doesn't change quickly

Disadvantage

the data we used could be out of date

sometimes the capital market is too changeable to be controlled and predicted

Conclusion

- Difficult to get detailed information of so many bonds.

- Quite a long time to finish running.

- Benefit > Cost, so it is not a bad choice to use it!

- Reasonable for problems and worth a try.

- Improve it according to the feedback.