BAN 432 Final Project

Group Members:

S184674 Meiqi Luo & S185438 Yu Chen

Seven Variables

- Sentiment Score GI
- Key Words in Context
- Investors' Passion
- Team Size
- Advisor Board
- Document Clustering
- Readability

Sentiment Score (GI)

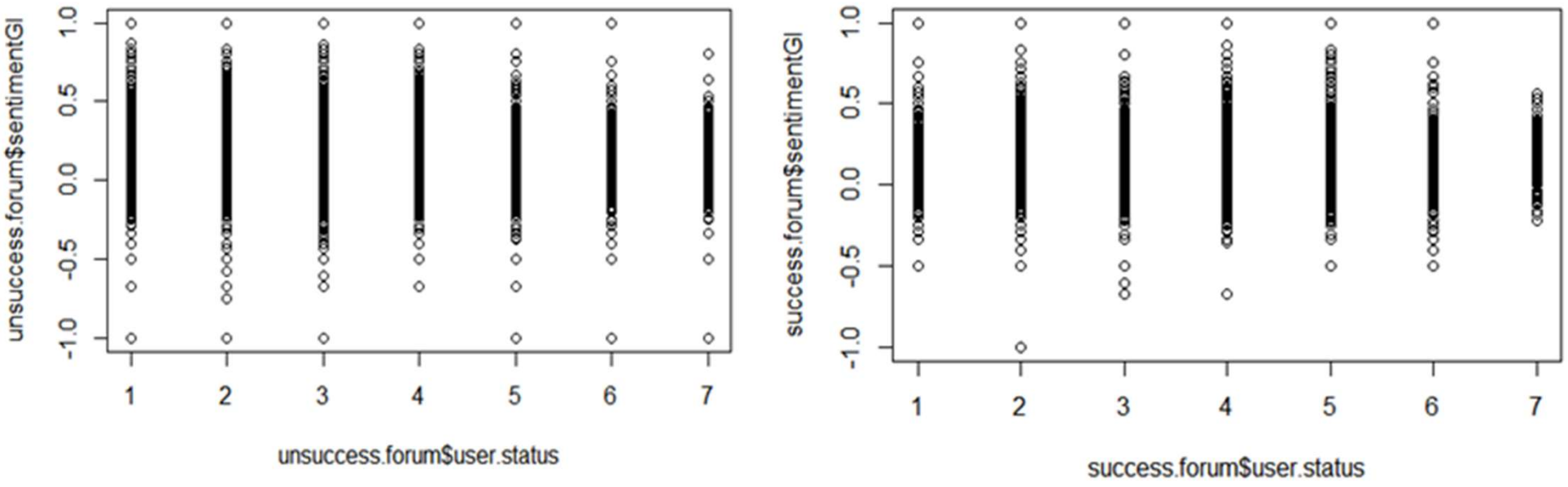
Hypothesis:

- Higher sentiment score stands for more optimistic the investors could be
- Indicate higher possibility to succeed

Sentiment Score (GI)

- Forum Texts
- Divided into two categories (success & failure)
- Firstly include all status users

Sentiment Score (GI)



Graph 1: Sentiment Score GI for users in all status

Sentiment Score (GI)

	Min.	Max.	Median	Mean	s.d.
Text from users of all status					
success	-1.0000	1.0000	0.1333	0.1449	0.1595
unsuccess	-1.0000	1.0000	0.1429	0.1591	0.1350
Text from users of status >= 4					
success	-0.6667	1.0000	0.1364	0.1473	0.1311
unsuccess	-1.0000	1.0000	0.1364	0.1460	0.1388
Text from users of status >= 5					
success	-0.5000	1.0000	0.1364	0.1453	0.1359
unsuccess	-1.0000	1.0000	0.1296	0.1369	0.1350
Text from users of status >= 6					
success	-0.5000	1.0000	0.1333	0.1415	0.1273
unsuccess	-1.0000	1.0000	0.1250	0.1330	0.1306

Table 2: Summary Statistics of Sentiment Score GI (Filtered Status)

Sentiment Score (GI)

- Exclude the texts posted by low-status users (only conclude users status >=5 make sense)

Sentiment Score (GI)

- t = 3.0962, p-value = 0.001969
- Significant difference at 95% confidence level
- Sentiment score GI can predict (statistically)
- Sentiment score > 0.13 more likely to succeed

Key Words in Context

Hypothesis:

- High risk & Lack of regulation
- Frequencies of keywords such as fraud and MLM
- Closely related to success or failure (Michiel 2018, para. 7)

Key Words in Context

- Forum Texts
- Keywords: scam, con, MLM, fraud, trick, deceive, lie, swindle, cheat
- Also excluded the users of lower than status 5
- Successful: 10 per 1000 words
- Unsuccessful: 11 per 1000 words
- Nonsignificant statistic difference (cannot predict)

Investors' Passion

Hypothesis:

- More frequently talked about
- More attention from potential investors
- More likely to succeed

Investors' Passion

- t = 1.9456 and p-value = 0.02837
- Means: successful (370) unsuccessful (250)
- Significant difference at 95% confidence level
- Investors' passion can predict (statistically)
- Freq of texts > 370 more likely to succeed

Team Size

Hypothesis:

- ICO with small team is less likely to succeed
- Call for joint effort of team members to make a project successful (Andrey 2018, para. 5)

Team Size

- Tibble ICO (Team Members)
- Count the team members numbers of both successful and unsuccessful ICOs

Team Size

	Min.	Max.	Median	Mean	s.d.
success	3	34	10	12	8
unsuccess	1	12	8	9	5

Table 4: Summary Statistics of Team Size (Both Success and Failure)

Team Size

- t = 2.2571, p-value = 0.0285
- Significant difference at 95% confidence level
- Team size can predict (statistically)
- Team size > 10 more likely to succeed
- Exception: small budget projects
- Size >= 10: 2792000; size < 10: 931500

Advisor Board

Hypothesis:

- Investors are more likely to invest an ICO project involving famous and experienced advisors
- More investors make a project more likely to success

Advisor Board

Our plan:

- Get a name list of famous ICO advisors
- Count the number of famous advisors in each ICO project
- Compare frequency table of successful and unsuccessful projects
- Conduct chi-square test

Advisor Board

	0	1	2	3	4	5	6
success	36	7	3	3	0	0	1
unsuccess	225	21	5	9	4	1	0

Frequency Distribution of Famous Name

Table 5: Frequency Distribution of Famous Name

p-value = 0.02636

Document Clustering(LDA)

Hypothesis:

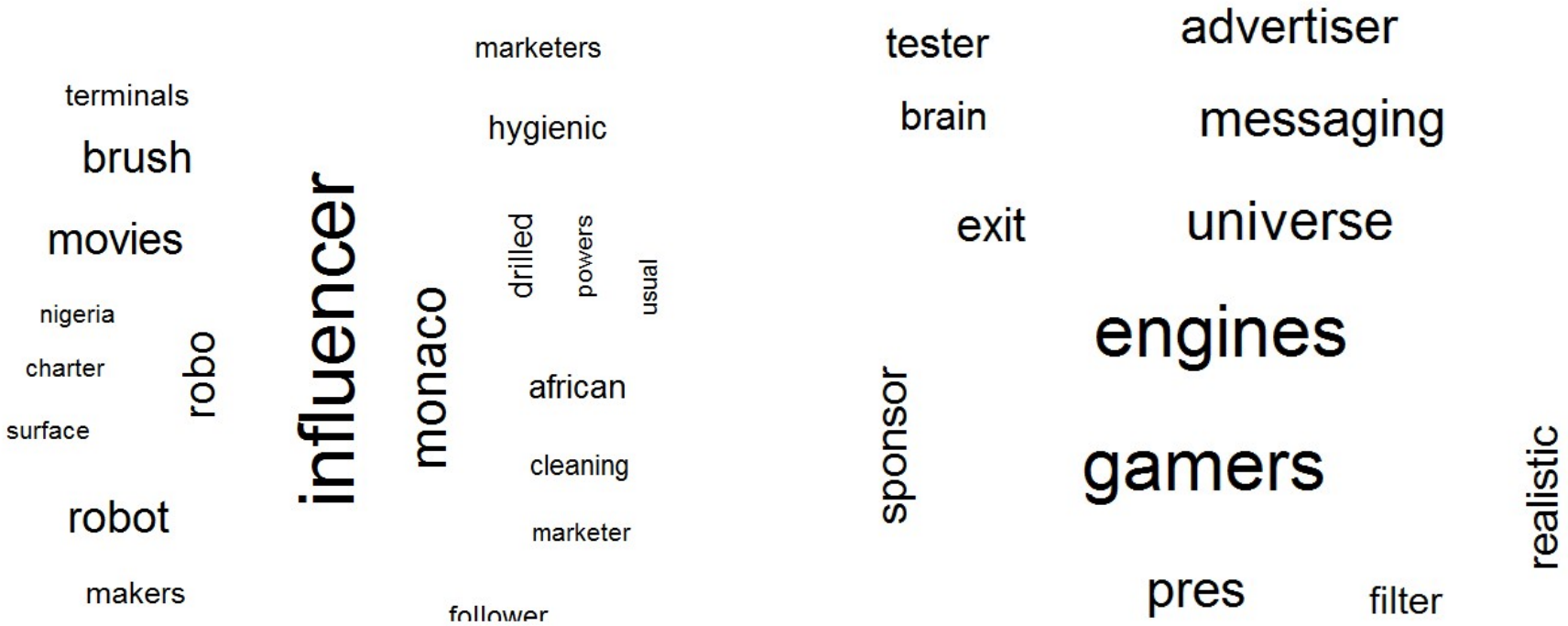
- Successful ICO project are more widespread in certain industries
- If investors invest money in ICO of these industries, they are more likely to make profit

Document Clustering(LDA)

Our plan:

- Remove ICO projects whose words are less than 300
- Use LDA to classify projects into 25 topics

Document Clustering(LDA)



Graph 2: Word Clouds for Topic 1 and Topic 2

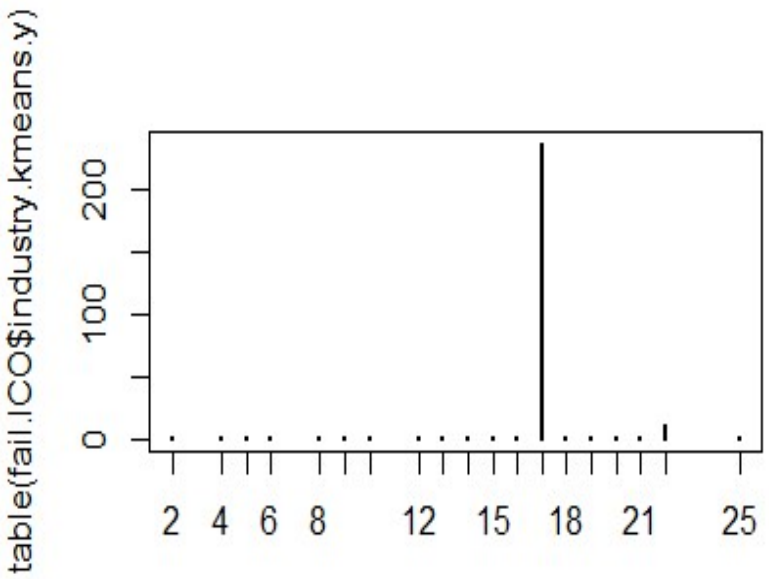
Document Clustering(LDA)

Document clustering based on whitepaper is ineffective:

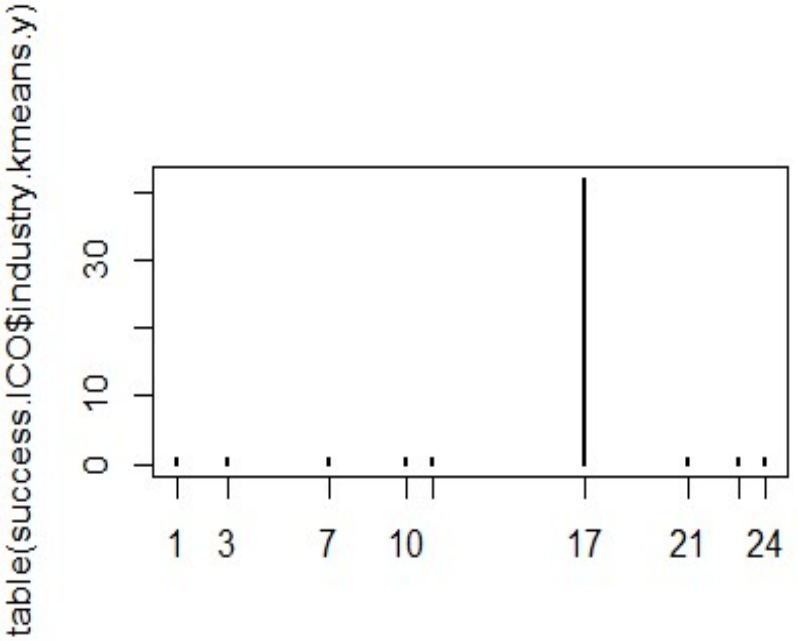
- Containing information about introduction, roadmap, team member profiles, funds allocation and so on
- Some ICO projects overlap industries
- Lack of regulated format

Document Clustering(K-means)

- To further proof our analysis in LDA part



Graph 3: Frequency of Each Unsuccessful ICO's assignment



Graph 4: Frequency of Each Successful ICO's assignment

WCSS = 954683

Readability

Hypothesis:

- Investors are more likely to invest in ICO projects whose whitepaper is relatively easy to read, since they can gather more information about this project

Readability

Our plan:

- Not choose variables related to complex words
- Use number of verbs, Fichtner's C, Yule's K and document length

Readability

Complex words are not complex for investors and professionals

Term	Frequency	syllable
biomedical	9	5
singularity	9	5
democratically	7	6
managerial	5	5
cooperative	4	5
evolutionary	4	6
appreciation	3	5
biological	3	5
inflationary	3	5
sophistication	3	5

Table 6 Complex Words in Whitepapers

Readability

Success	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Verbs	182	1601	2050	2559	3386	7537
Fichtner's C	6.82	75.75	103.97	100.7	129.32	246.32
Yule's K	38.23	79.01	87.53	100.02	98.15	297.66
Doc Length	7.78	8.61	8.87	8.95	9.34	10.31

Table 7: readability statistics for success firms

Unsuccess	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Verbs	10	1119	1791	2024	2618	13285
Fichtner's C	0.15	80.82	105.07	108.14	134.63	424.04
Yule's K	37.92	77.52	89.11	99.28	103.89	1167.95
Doc Length	6.34	8.26	8.74	8.67	9.15	10.771

Table 8: readability statistics for fail firms

Readability

Average number of verb and length of success firms are greater than unsuccessful firms

- ICO projects are funding for start-ups
- Besides whitepapers, there are only few official information concerning this ICO project
- Document length :0.004125
- Number of verbs :0.02467

Conclusion

Statistically have prediction power:

- Key Words in Context
- Investors' Passion
- Team Size
- Advisor Board
- Readability

Conclusion

- In reality, these variables have limited prediction power
- Be very careful
- Make an overall evaluation of all available information