Introduction to Data Mining

Scott Powers and Ryan Wang

December 21, 2013

RMS Titanic

- In 1912, the Titanic sank after colliding with an iceberg
- Total number of passengers and crew: 2224
  - How many people survived?

RMS Titanic

- In 1912, the Titanic sank after colliding with an iceberg
- Total number of passengers and crew: 2224
    - How many people survived? 722
- Kaggle: Can we predict which passengers survived?

The Data

train.csv:

ID	Su	C	Name	S	Age	Sib/Sp	Par/Ch	Fare	Cabin	Port
1	0	3	Braund, Owen	M	22	1	0	7.25		S
2	1	1	Cumings, Florence	F	38	1	0	71.28	C85	C
3	1	3	Heikkinen, Laina	F	26	0	0	7.93		S
4	1	1	Futrelle, Lily	F	35	1	0	53.10	C123	S
5	0	3	Allen, William	M	35	0	0	8.05		S
6	0	3	Moran, James	M	NA	0	0	8.46		Q
...	...	...	...	...	...	...	...	...	...	...
891	0	3	Dooley, Patrick	M	32	0	0	7.75		Q

test.csv:

ID	Su	C	Name	S	Age	Sib/Sp	Par/Ch	Fare	Cabin	Port
892	?	3	Kelly, James	M	35	0	0	7.83		Q
893	?	3	Wilkes, Ellen	F	47	1	0	7.00		S
894	?	2	Myles, Thomas	M	62	0	0	9.69		Q
...	...	...	...	...	...	...	...	...	...	...
1309	?	3	Peter, Michael	M	NA	1	1	22.36		C

Brainstorm: What kind of person is more likely to survive?