

Titanic Naive Bayes Model

Joshua Kristanto, Jose de la Rosa, Albert Eng
Supervised by Carol Shubin
California State University Northridge

Original Dataset

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
10	11	1	3	Sandstrom, Miss. ...	female	4.0	1	1	PP 9549	16.7000	G6	S

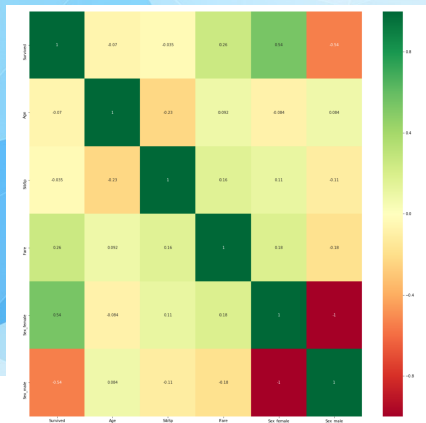
Dataset after Cleaning

	Survived	Sex	Age	SibSp	Fare
0	0	male	22.0	1	7.2500
1	1	female	38.0	1	71.2833
2	1	female	26.0	0	7.9250
3	1	female	35.0	1	53.1000
4	0	male	35.0	0	8.0500
5	0	male	NaN	0	8.4583
6	0	male	54.0	0	51.8625
7	0	male	2.0	3	21.0750
8	1	female	27.0	0	11.1333
9	1	female	14.0	1	30.0708
10	1	female	4.0	1	16.7000
11	1	female	58.0	0	26.5500
12	0	male	20.0	0	8.0500
13	0	male	39.0	1	31.2750
14	0	female	14.0	0	7.8542
15	1	female	55.0	0	16.0000
16	0	male	2.0	4	29.1250
17	1	male	NaN	0	13.0000
18	0	female	31.0	1	18.0000
19	1	female	NaN	0	7.2250
20	0	male	35.0	0	26.0000
21	1	male	34.0	0	13.0000
22	1	female	15.0	0	8.0292
23	1	male	28.0	0	35.5000
24	0	female	8.0	3	21.0750
25	1	female	38.0	1	31.3875
26	0	male	NaN	0	7.2250

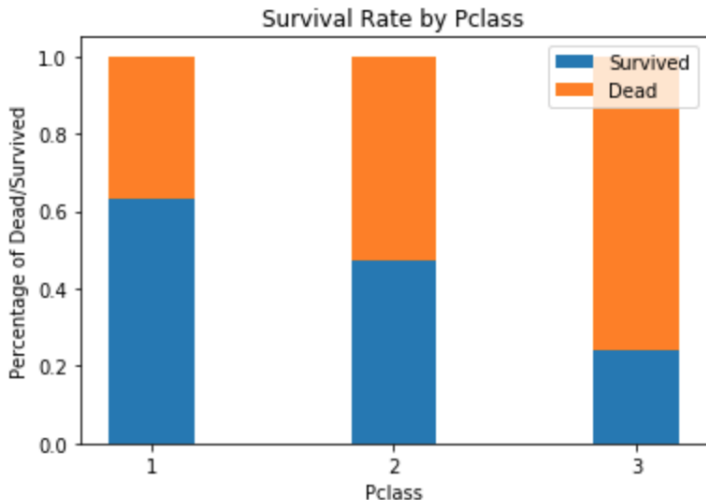
Obtaining Dummies for Gender

	Survived	Age	SibSp	Fare	Sex_female
0	0	22.000000	1	7.2500	0
1	1	38.000000	1	71.2833	1
2	1	26.000000	0	7.9250	1
3	1	35.000000	1	53.1000	1
4	0	35.000000	0	8.0500	0
5	0	29.699118	0	8.4583	0
6	0	54.000000	0	51.8625	0
7	0	2.000000	3	21.0750	0
8	1	27.000000	0	11.1333	1
9	1	14.000000	1	30.0708	1
10	1	4.000000	1	16.7000	1
11	1	58.000000	0	26.5500	1
12	0	20.000000	0	8.0500	0
13	0	39.000000	1	31.2750	0
14	0	14.000000	0	7.8542	1
15	1	55.000000	0	16.0000	1
16	0	2.000000	4	29.1250	0
17	1	29.699118	0	13.0000	0
18	0	31.000000	1	18.0000	1
19	1	29.699118	0	7.2250	1
20	0	35.000000	0	26.0000	0
21	1	34.000000	0	13.0000	0
22	1	15.000000	0	8.0292	1
23	1	28.000000	0	35.5000	0
24	0	8.000000	3	21.0750	1
25	1	38.000000	1	31.3875	1
26	0	29.699118	0	7.2250	0
27	0	19.000000	3	263.0000	0

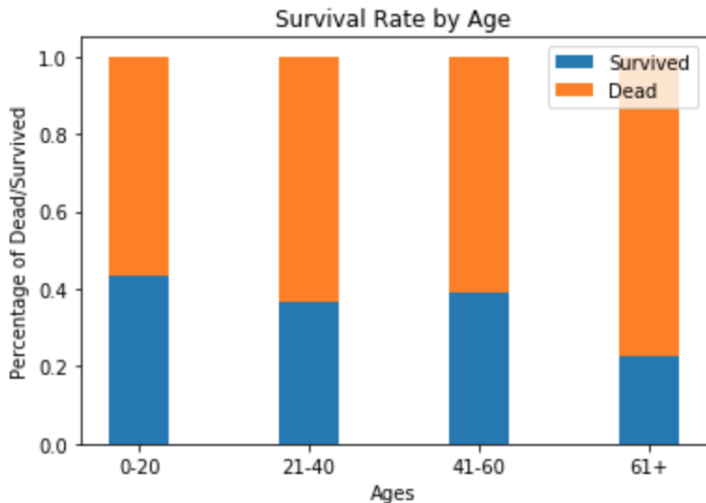
Heatmap of Dataset



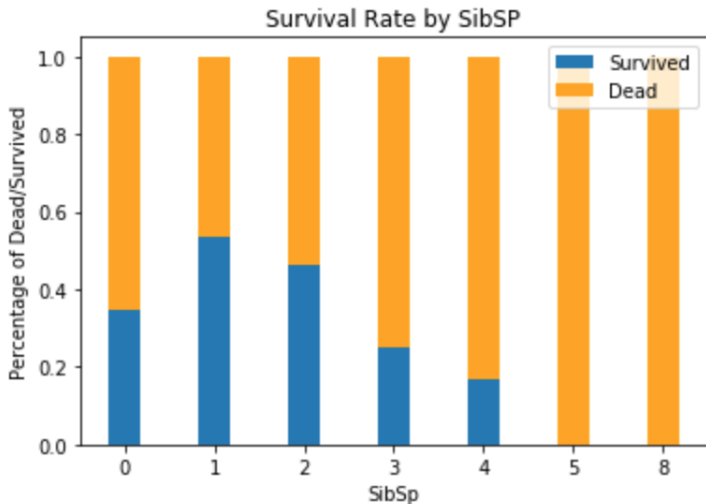
Survival vs Pclass Histogram



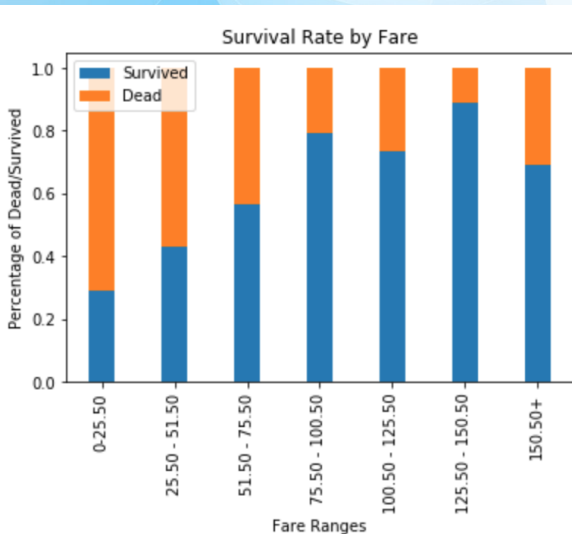
Survival vs Age Histogram



Survival vs Sibling/Spouse Histogram



It's Not Fare



Conclusion

7819

JokeKi



0.76555

4

1h