

FORECASTING SURVIVAL RATE USING KAGGEL TITANIC DATASET

NAVEEN BALARAJU
UNIVERSITY AT BUFFALO

GOAL: Forecasting survival rates for Titanic dataset

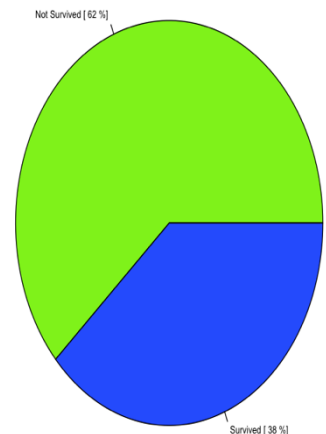
FEATURE SELECTION & DATA PREPROCESSING:

Here, I'm assuming Name of the person as the unique identity of an individual and have selected features that are necessary for forecasting the survival rates. The Age variable contains 177 missing values. So, I've replaced missing values with the median value of Age i.e. 28 and have Binned the variables into different levels.

EXPLORATORY DATA ANALYSIS:

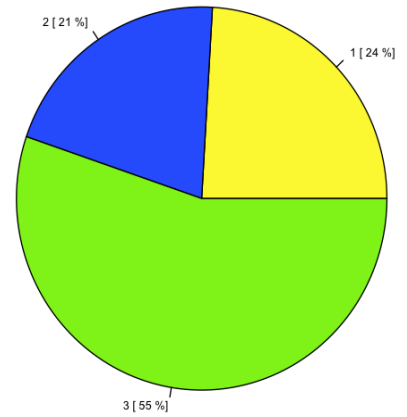
1. Analysis of Survived variable - From the pie chart, it shows that only 38% of the people aboard survived in Titanic sinking incident.

PIE CHART FOR SURVIVAL RATE,[0:Not Survived, 1:Survived]



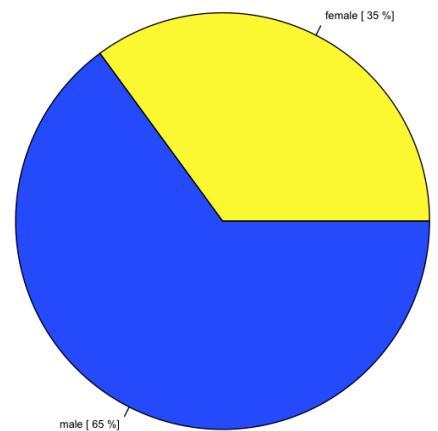
2. Analysis of variable Pclass - The pie chart shows that there are 55% of people were in class 3, 21% in class 2 and 24% in class 1.

PIE CHART FOR TICKET CLASS



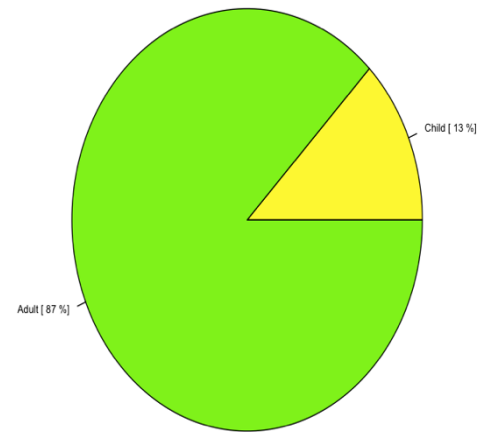
3. Analysis of variable sex -The data contains 65% male population and 35% female population.

PIE CHART FOR GENDER



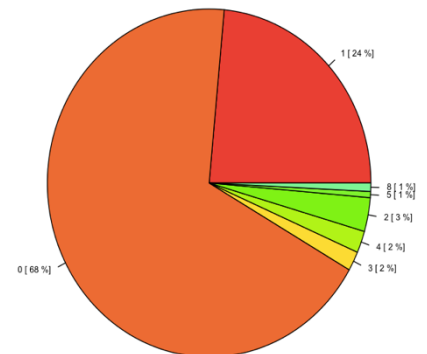
4. Analysis of variable Age- For the further analysis, I've considered all the people of age below 18 as "Children" and of age 18 and above as "Adults". The data contains 113 people with age below 18 i.e. 13% of the population on board of Titanic are children and rest 87% are Adults.

PIE CHART FOR AGE GROUP



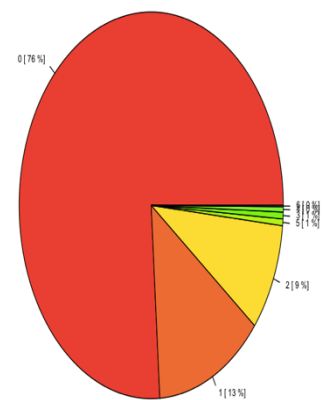
5. Analysis of variable Sibsp - The pie chart shows that 68% of the people were travelling alone. While, 23% were aboard with 1 sibling/Spouse and rest 9% were aboard the Titanic with 2 siblings or more.

PIE CHART FOR NUMBER OF SIBLINGS/SPOUSES ABOARD THE TITANIC



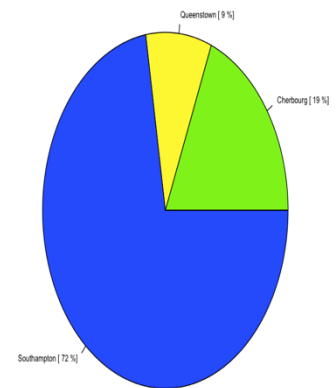
6. Analysis of variable Parch - The pie chart shows that 76% of the people aboard the Titanic were travelling alone or with spouse. Whilst, 13% aboard were with one parent/child and rest 9% were traveling with more than one parents/children.

PIE CHART FOR NUMBER OF PARENTS/CHILDREN ABOARD THE TITANIC



7. Analysis of variable Embarked - The pie chart shows that 72% of the people had port of departure as Cherbourg, 19% Queenstown and rest 9% had the port of departure as Southampton.

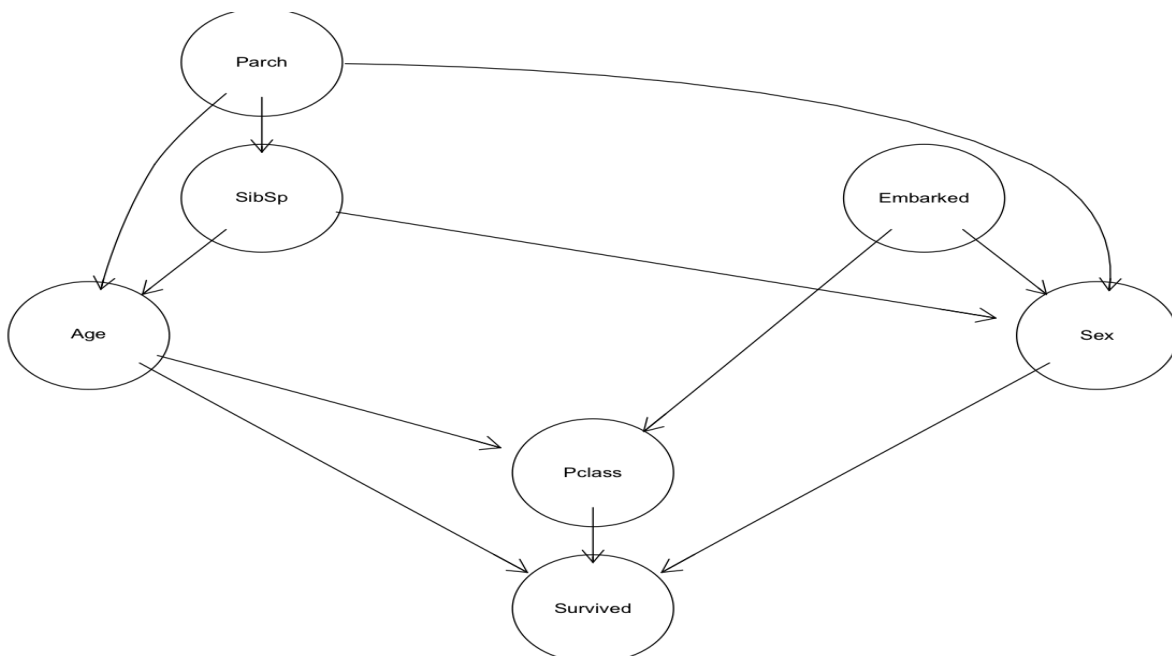
PIE CHART FOR PORT OF EMBARKATION(C:Cherbourg,Q:Queenstown,S:Southampton)



BAYESIAN NETWORK FOR FORECASTING SURVIVAL RATES:

To find the optimal Bayesian network, I've have used Max-min Hill climbing algorithm with "k2" as the score.

Optimal DAG:



CONCLUSIONS:

1. Female and adult's belonging to pclass-1 and pclass-2 survived with the probability of 0.97 and 0.90 respectively. While, Female adults pclass-3 did not survive

, , Sex = female, Age = Adult

		Pclass		
Survived		1	2	3
Not Survived		0.02428147	0.09427828	0.513751
Survived		0.97571853	0.90572172	0.486249

2. Most of the children and females seated in the pclass1, pclass2 and pclass3 survived with the probability of 0.87,0.99,0.54 respectively. While, the probability of child and male in pclass1 and pclass2 survived with probability of 0.98 and 0.81 respectively, while children who are male belonging to pclass 3 did not survive.

, , Sex = female, Age = Child

		Pclass		
Survived		1	2	3
Not Survived		0.128866	0.003448276	0.4572447
Survived		0.871134	0.996551724	0.5427553

, , Sex = male, Age = Child

		Pclass		
Survived		1	2	3
Not Survived		0.01020408	0.1842105	0.7669246
Survived		0.98979592	0.8157895	0.2330754

3. The results also shows that Adult males did not survive as their probability of being dead is high for all the three classes:

pclass1 death rate - 65.7%

pclass2 death rate - 91.7%

pclass3 death rate - 87.8%

Overall, most of the adult male did not survive.

, , Sex = male, Age = Adult

		Pclass		
Survived		1	2	3
Not Survived		0.6524347	0.91716738	0.8781858
Survived		0.3475653	0.08283262	0.1218142

4. All the above findings gives an intuition that most of the Women and children were evacuated first. Although, child(male) and adult(female) belonging to pclass did not survive.
5. Female adult belonging to pclass1 - SURVIVED
6. Male adult belonging to pclass3- DID NOT SURVIVE.