Data Analysis Assignment (Titanic Dataset)

S Sai Girish

Overview :

Using Machine Learning to create a model that predicts which passengers survived the Titanic shipwreck.

Data :

Using kaggle Titanic Dataset (https://www.kaggle.com/c/titanic/data) and try to understand and analyze data.

Analysis:

- There are 891 records.
- Columns : PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked.
- We fnd that 74.2% of all the females survived compared to 18.8% of all the males.
- We find that 62.9% of all the high class, 47.2% of all the middle class and 24.2% of all the lower class people survived.
- We find that 46.6% of all the people who have siblings/spouses survived compared to 34.5% of those people who didn't have any siblings/spouses.
- We find that 51.1% of all the people who had parents/children survived compared to 34.3% of all the people who didn't have parents/children.
- We also find that the data is not complete, there are few missing values. Age column has 177 missing values, cabin column has 687 missing values and Embarked column has 2 missing values.

Conclusion:

We can conclude that chances of surviving are high for females, high class, people who have siblings/spouses, people who have parents/children when compared to males, low class, people who don't have siblings/spouses and people who don't have parents/children.