**PROGRAMMING ASSIGNMENT 2 – Ankit Nashine**

**1) Using google collab. Installing tensorflow. Version 2.4.1**

Text

Description automatically generated

**Installing Libraries**

Text, letter

Description automatically generated

**2) Loading the full data for Titanic (will split it into test and train later)**

Graphical user interface

Description automatically generated with medium confidence

**3) Exploratory analysis of the data**

Table

Description automatically generated

**Graphs outlining nature of attributes**

Chart, bar chart, histogram

Description automatically generatedChart, bar chart

Description automatically generated Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

**4) Data cleaning (created a new dataframe df with only those columns that are useful for analysis and dropping null rows)**

Graphical user interface, table

Description automatically generated with medium confidence

Table

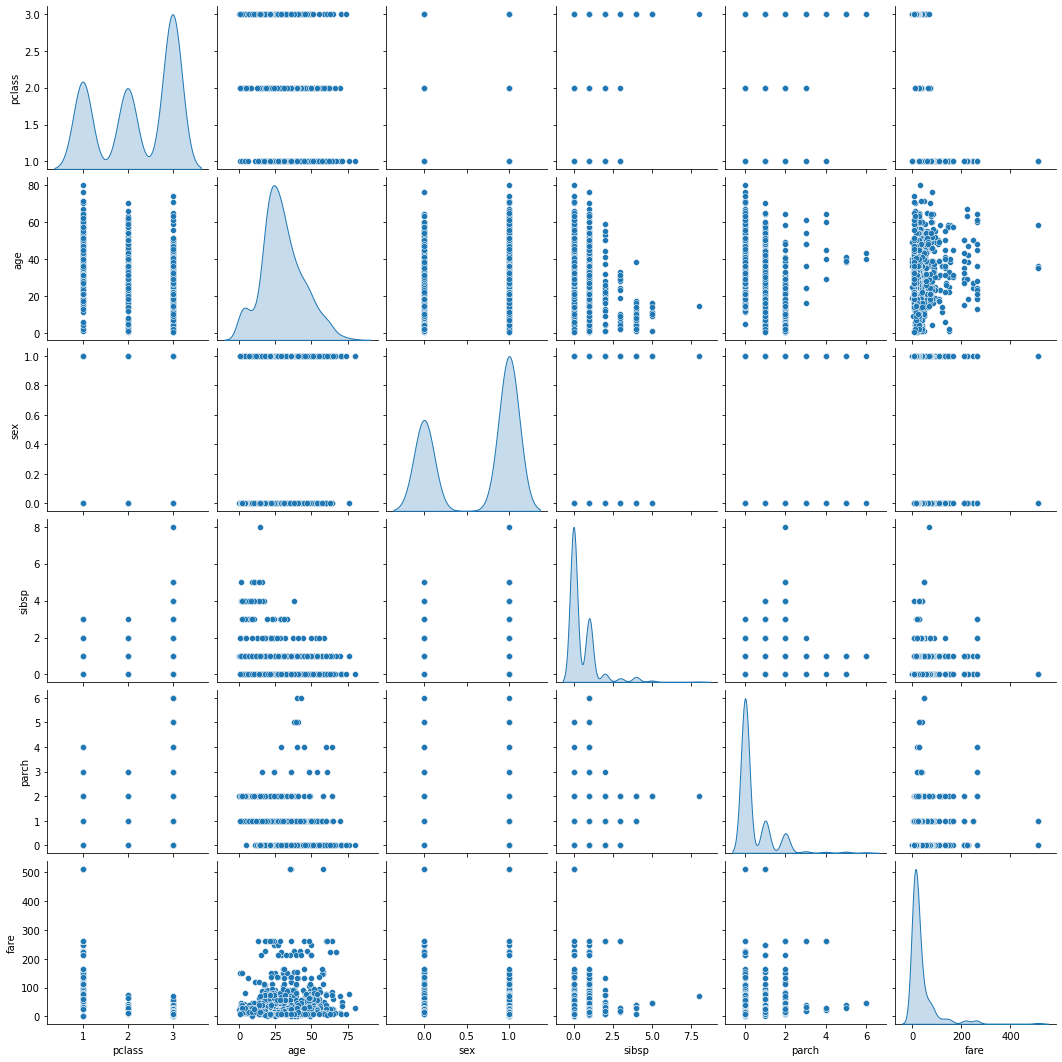
Description automatically generated

**Creating dummies for embarked and changing datatype**

A picture containing text

Description automatically generated

**Pairplot to show the relationship for combination of variable**



**6) Visualizations with finalized data (df)**

People with higher fare have more survival rate.

A picture containing histogram

Description automatically generated

Children under 10 have better survival rate and people aged between 20-30 have highest death rate.

Chart, histogram

Description automatically generated

Parch 1, 2, 3 have higher survival rate and others have higher death rate (0 being highest)

Chart, bar chart

Description automatically generated

sibsp 1 have higher survival rate and others have higher death rate.

Chart, bar chart

Description automatically generated

Embarked C have higher survival rate, others have high death rate.

Chart, bar chart

Description automatically generated

Females have higher survival rate whereas men have higher death rate.

Chart, bar chart

Description automatically generated

Pclass 1 passengers have high survival rate whereas class 3 have lowest survival rate.

Chart, bar chart

Description automatically generated

**Partitioning data for training and testing.**

Table

Description automatically generated

**Preprocessing data to use in models**

Table

Description automatically generated

Table

Description automatically generated

**5) Try few different models.**

* Creating and compiling tensorflow keras ANN.

Table

Description automatically generated

* Training Neural Network model and checking run time of model.

Table

Description automatically generated

Model accuracy is 81.68% and runtime is 52.39 seconds.

**Checking accuracy on test set.**

Graphical user interface, text, application

Description automatically generated

**Other models building**

Graphical user interface, text, application

Description automatically generated

A picture containing table

Description automatically generated

**Training accuracy of other models and time.**

Text

Description automatically generated with medium confidence

**Testing accuracy**

Graphical user interface, text, application

Description automatically generated

**Variable importance**

Graphical user interface

Description automatically generated

**6. Would you {and an assumed passenger e.g Woman / 30 year old} survived the trip?**

Chart, bar chart

Description automatically generated Chart, histogram

Description automatically generated

Based on the graphs above a woman has a high chance of survival whereas 30-year-old has less chance of survival. Based on variable importance chart a page before, age had more importance so considering age, the chance of survival is slightly less than death chance.

Predicting my and jack rose survival (assuming values for my) using ANN and Random Forest (as it has good testing and training accuracy among all)

Graphical user interface, text, application

Description automatically generated

This model gives me just 3.9% survival chance whereas for jack its 12.7% and rose its 91.8%.

**7. Short write-up [1 page] of you findings in layman terms... who survives .. who dies.**

Based on visualizations in { 6 - Visualizations with finalized data (df) }

a) it is very clear that females are more likely to survive and men are more likely to die.

b) based on fare, high fare passengers (>50) are more likely to survive and less fare passengers are more likely to die.

c) People with age less than 10 and greater than 50 are more likely to survive and between 10 and 50 are more likely to die.

d) Parch 1, 2, 3 have higher survival rate and others have higher death rate (0 being highest).

e) sibsp 1 have higher survival rate and others have higher death rate.

f) Embarked C have higher survival rate, others (s, Q) have high death rate.

g) class 1 passengers have high survival rate whereas class 3 have highest death rate.

**Which combination of attributes is best?**

Old – Female – Embarked C – Class 1 passenger/high ticket fare – sibsp 1 – parch 1