

# Project

## Course: Big Data & Data Analytics - II

Intern Id: 180022

Intern's Name: Shubham Chand

Period: 10-10-2021

### Project Activity 3: Predicting the Survival of Titanic Passengers

#### Section - A

Answers of all the questions (Q1 to Q9) are uploaded on github whose link is given below:

[https://github.com/shubhamchand03/180022\\_W10\\_BigData\\_PA3](https://github.com/shubhamchand03/180022_W10_BigData_PA3)

#### **Q10) What conclusions are you derived from the analysis?**

- From the analysis People with higher socioeconomic class had a higher rate of survival.
- Females had Higher chance of survival compared to males.
- People who are having no siblings or spouses are less to likely to survive than those with one or two.
- People traveling with 1-3 parents or children have more chance of surviving compared to who are travelling solo.
- Kids are more likely to survive than any other age group.
- People with a recorded Cabin number are, more likely to survive.

#### Section – B

#### **Q1) Which is the best model?**

- **K-nearest neighbor is the best model:**
  1. Simple & intuitive — The algorithm is very easy to understand and implement
  2. Memory based approach — Allows it to immediately adapt to new training data
  3. Variety of distance metrics — There is flexibility from the users side to use a distance metric which is best suited for their application (Euclidean, Minkowski, Manhattan distance etc.)