PUNE INSTITUTE OF COMPUTER TECHNOLOGY,
DHANKAWADI PUNE-43.

A Mini-Project Report
On

Titanic Survivor Classification

SUBMITTED BY

NAME:
Saloni Mittal
(41251)
Vidula Magdum
(41248)
Ashwini Kurhe
(41245)

CLASS: BE-II

COMPUTER ENGINEERING DEPARTMENT
Academic Year: 2022-23

PUNE INSTITUTE OF COMPUTER TECHNOLOGY,
DHANKAWADI PUNE-43.

CERTIFICATE

This is to certify that ***Saloni Mittal (41251)*** student of B.E. (Computer Engineer- ing Department) Batch 2022-2023, have satisfactorily completed a report on "**Titanic Survivor Classification**" towards the partial ful- fillment of the fourth year Computer Engineering Semester VII of SPPU.

Dr. G. V. Kale
Head of Department,
Computer
Engineering

Date:

Place:

Titanic Survivor Classification

Contents

List of Figures

1 PROBLEM STATEMENT :

Build a machine learning model that predicts the type of people who survived the Titanic shipwreck using passenger data (i.e. name, age, gender, socio-economic class, etc.).

2 ABSTRACT :

The titanic dataset is one of the most widely used datasets for learning the fundamentals of machine learning. It contains information on all of the passengers aboard the RMS Titanic, which was tragically lost at sea.

The "unsinkable" RMS Titanic sank on April 15, 1912, during her maiden voyage, after colliding with an iceberg. Unfortunately, there were insufficient lifeboats to accommodate everyone onboard, resulting in the deaths of 1502 of the 2224 passengers and crew. While survival required some luck, it appears that some groups of people were more likely to survive than others.

3 REQUIREMENTS :

Hardware Requirements :-

a. 64 bit OS
b. 16 GB RAM

Software Requirements :-

a. Python 3

b. Jupyter Notebook/ Google Colaboratory

4 INTRODUCTION :

This particular problem statement involved predicting those passengers who were likely to survive the sinking of the Titanic. The dataset contains some details about the passengers, like their age, fare and class. The data has categorical features as well as some correlated features. The initial stage of the project involved converting some features into their numerical form and handling of missing values as well. Various visualization techniques were used to represent the data, in order to understand the patterns better. A logistic regression model was trained on the data to give commendable accuracy scores.

0.1 Logistic Regression

Based on a set of independent variables, logistic regression calculates the likelihood of an event occurring, such as voting or not voting. Because the outcome is a probability, the dependent variable has a range of 0 to 1. A logit transformation is applied to the odds in logistic regression, which is the probability of success divided by the probability of failure. This is also referred to as log odds or the natural logarithm of odds.

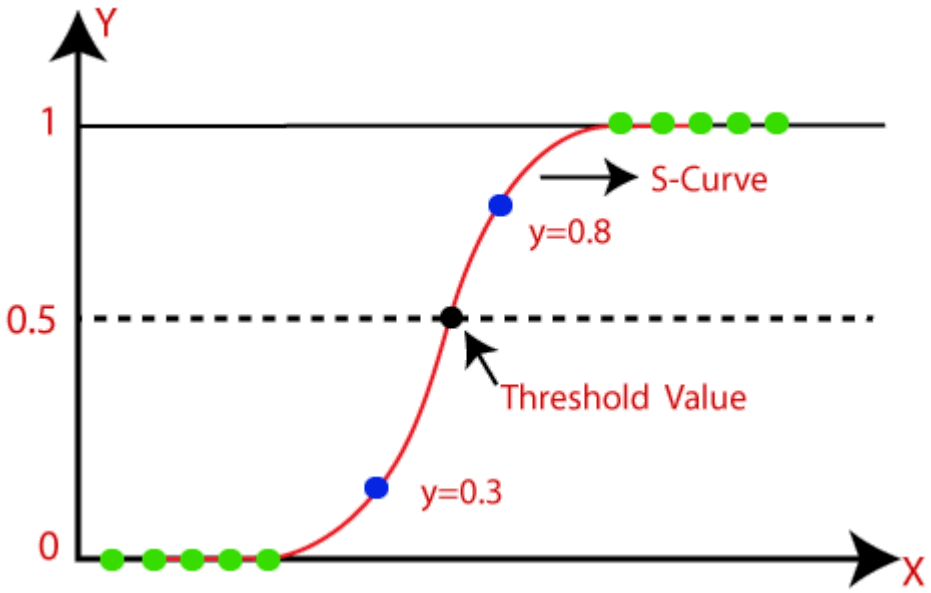

Figure 1: Logistic Regression

5 OBJECTIVE :

Building a machine learning model to predict whether a person survived the sinking of the Titanic or not, based on the given features.

6 SCOPE :

- The dataset consisted of the following features -

```
survived          int64
pclass            int64
sex              object
age             float64
sibsp             int64
parch             int64
fare            float64
embarked         object
class          category
who              object
adult_male         bool
deck           category
embark_town      object
alive            object
alone              bool
```

- The dataset had 891 data points to train on
- It was split into train and test sets
- The following accuracy metrics were used to evaluate the classifier
    - recall
    - precision
    - f1 measure
    - accuracy

7 RESULT

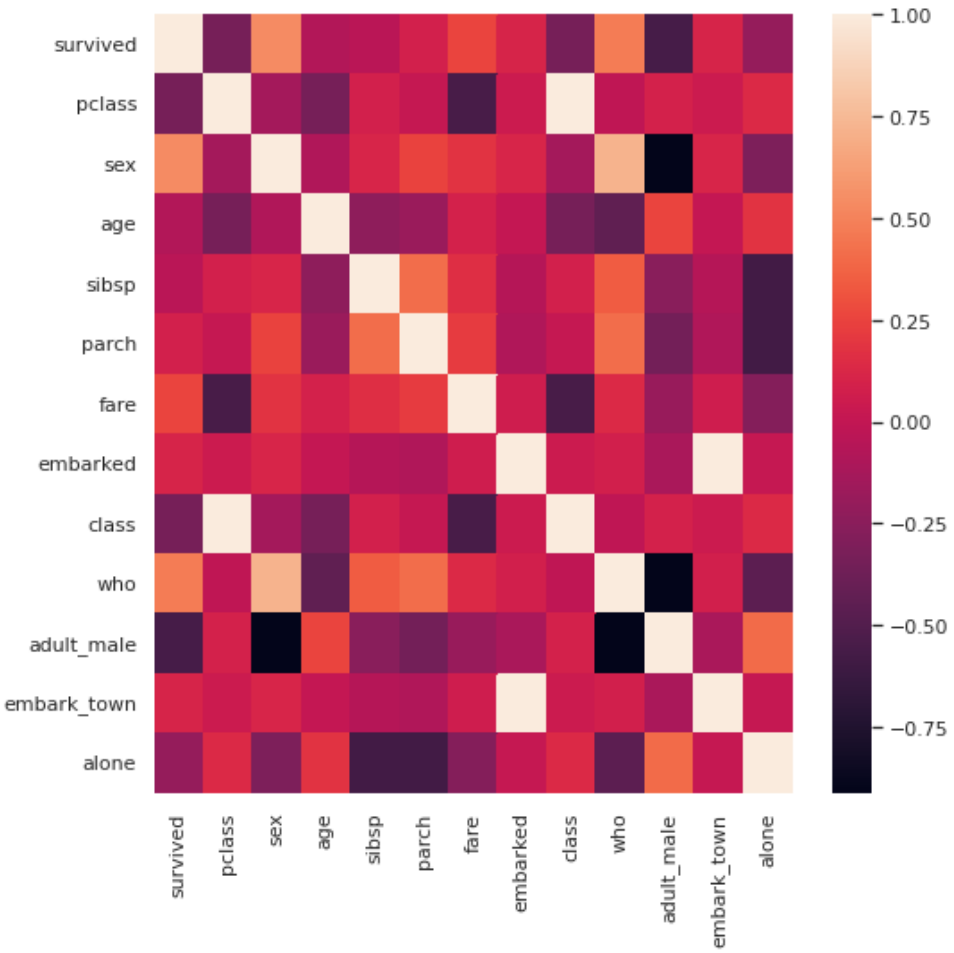

Figure 2: Correlation matrix

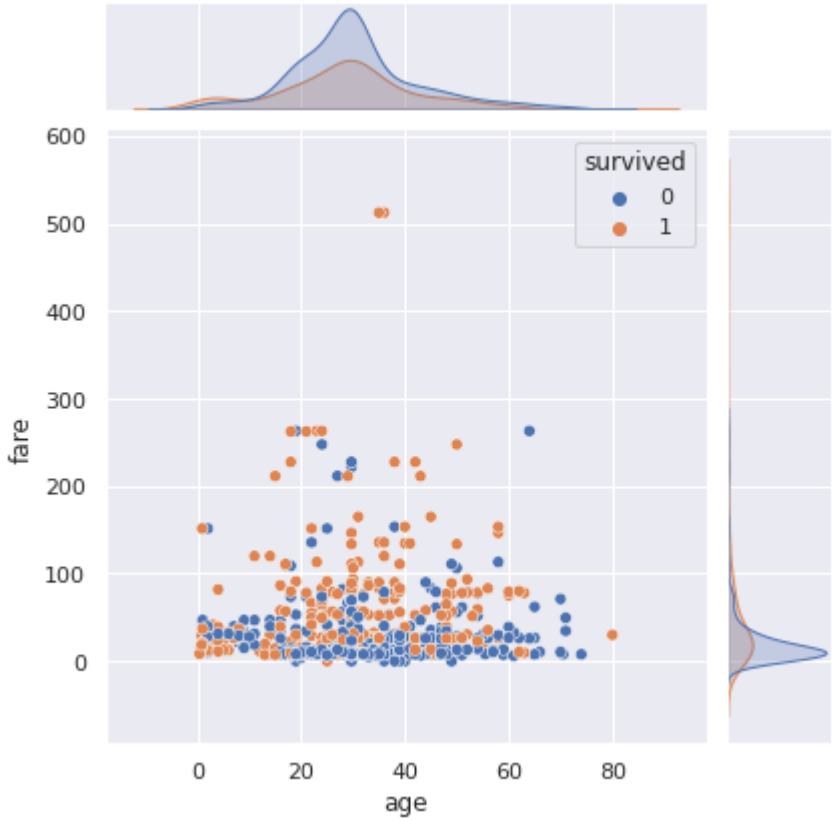

Figure 3: Joint plot

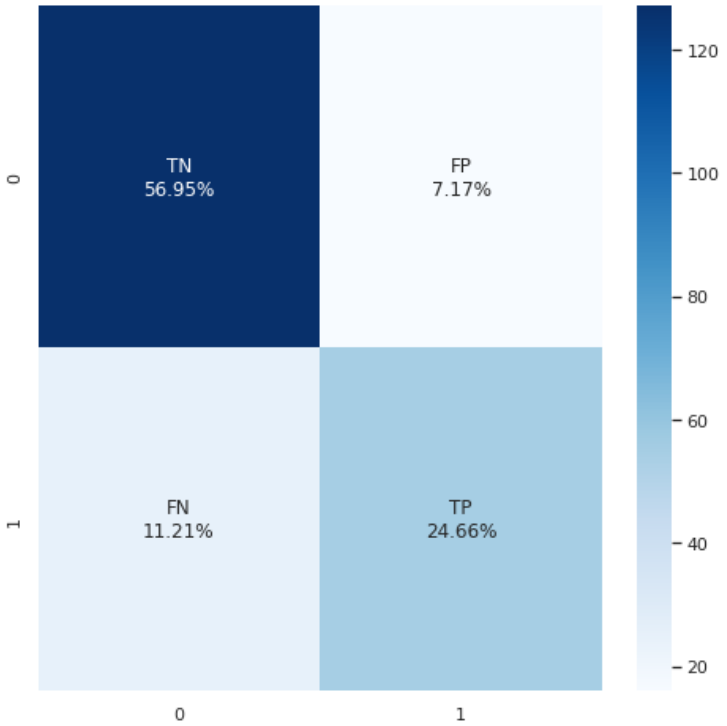

Figure 4: Confusion matrix

8 CONCLUSION :

Hence, we trained a logistic regression model on the titanic dataset to obtain the following accuracy scores -

- Accuracy score - 0.816
- Recall score - 0.687
- Precision score - 0.774
- F1-score - 0.728

9 REFERENCES :

[1] https://www.kaggle.com/c/titanic.

[2] https://scikit-learn.org/stable/

[3] https://www.ibm.com/in-en/topics/logistic-regression