# Week 4 – Flask Deployment Assignment Report

Problem Statement

Determining the characteristics of survivors from the Titanic crash.

Given the provided dataset, is it possible to determine the survivors based on key features?

Data Acquisition & Preparation

Data obtained from: <https://www.kaggle.com/c/titanic/data?select=test.csv>

Dataset is divided into:

*Test* data – Contains all features necessary for model training.

*Train* data – Contains all features in the test.csv, apart from ‘Survived’, which forms the feature to be predicted. This will similarly be the predicted value for the model to be deployed.

Data analytics

Data Intake Report

Name: Deployment on Flask – Titanic Survival Predictor

Report date: 17th October 2021

Internship Batch: <Enter your batch code from Canvas course>

Version: 1.0

Data intake by: Teddy Waweru

Data intake reviewer:

Data storage location:

**Tabular data details:**

|  |  |
| --- | --- |
| **Total number of observations** | 891 |
| **Total number of files** | 2 |
| **Total number of features** | 12 |
| **Base format of the file** | .csv |
| **Size of the data** | 87.7KB |

**Note: Replicate same table with file name if you have more than one file.**

**Proposed Approach:**

* Mention approach of dedup validation (identification)
* Mention your assumptions (if you assume any other thing for data quality analysis)

Model creation

Model training

Model evaluation

Model Deployment

Results & Conclusion