

Group Project / Assignment - Milestone 1

Background

The following scenario comes from a former business client of DMC, and serves as a “training / testing” of each team’s capabilities before handling the actual scenario in later milestones (e.g. MS_2).

The name of the client is confidential, however, it is in the business of “marine accident insurance”, specifically on passenger / leisure ships like cruises. Over many decades, the client has processed claims from thousands of cases, and consolidated a lot of passenger data.

DMC will be passing to you 2 files :

- MS_1_Scenario_train.csv (for your EDA & training/building ML)
- MS_1_Scenario_test.csv (for testing your ML Model, and evaluation)

In the event of accidents in the high seas, whether the ship is damaged or entirely destroyed, lives of women, men, children will be lost. Of those who survived the accident, many will sustain all sorts of physical / mental injuries. The insurance claims are the highest for passenger who lost their lives.

The files that DMC passed to you, contains data like age or gender of passenger, the cabin they stayed, and most importantly, whether they survived the accident or not. Do note that all sensitive “identification data” like passenger’s name, has been painstakingly replaced with fictitious-yet-realistic names. (In the event of security leak, the damage will be contained to both DMC and its client).

The next section describes the set of tasks / instructions for this MS_1 project.

Task Instructions / Info

A) Based on the raw data in the training set, please perform analysis and help the client answer the following questions :

(i) List all the important characteristics of passengers who survived the accident

(ii) List all the important characteristics of passengers who died in the accident

These has important implications on whether an insurance company should / should not sell insurance to certain “group of individuals”, and for those who can be insured, what is the amount of premium to be charged?

B) The client also requested a “program” from your team, which could help them to decide whether certain individuals are eligible to be insured or not. (e.g. if an individual is likely to survive a maritime accident, then it could be eligible for insurance). It would be ideal if the program could be easily configured and integrated into the client’s existing IT infrastructure and OS platform.

C) While there are many other evaluation metrics for a trained ML model, your team should, at the very least, display the following information (if not more) on the model’s performance :

- The “usual” error metrics (e.g. MAE, MSE, RMSE, etc)
- Confusion Matrix, Precision, Recall, F1, etc

Please provide a summary & explanation of your ML Model’s performance, don’t just type out / paste a screen capture of the metrics, without any explanation!

D) As a team, please do some self-reflection on your system’s performance, and provide a few concluding paragraphs in your report that answers the following questions :

- Can the client place complete (100%) faith / trust in your ML model’s ability to help them achieve their objectives?
- If not, what is the estimated %-tage of trust client should place on your model’s accuracy? (i.e. 90%? 80%? < 50%??, etc)
- Regardless of 100% or some lesser percentage, please provide evidence / reasons to justify your percentage figures!

E) With regards to the context / meaning of each column in the data files, please refer to **Appendix A**, which provides the necessary description that applies to both *.CSV files (i.e. for training and for testing purposes)

APPENDIX A

(Passenger Data Description)

The following table gives the names of each column, as well as a short description of its data, possible values, data type, etc.

| Column Name | Data Type | Description |
|---------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Passenger ID | Number | A running number that serves as a unique id to locate each row of record |
| Passenger Fare | Text | Each cell value is pre-fix with '\$' symbol, followed by a decimal value. (E.g. \$27.3583). Some processing may be required ... |
| Ticket Class | Number | A "rough indicator" of passenger's wealth 1 == 1 st class ticket bought by passenger 2 == 2 nd class ticket, and so forth ... |
| Cabin | Text | The unique ID identifying the passenger's accommodation aboard the ship |
| Embarkation Country | Alphabet | The country's name is deemed "sensitive", so has been de-sensitized. (e.g. 'S' could mean any country whose name starts with it, like Spain, Singapore, Slovenia, ...) |
| Name | Text | Passenger's name is deemed "sensitive", so has been replaced with fictitious, yet seemingly realistic values |
| Age | Number | Whole number, age at which the passenger boarded the ship |
| Gender | Text | Text data, the sex of the passenger at the point he/she/.. boarded the ship (hey, I'm being open-minded ...) |
| NumSiblingSpouse | Number | Whole number, the sum of the individuals who : <ul style="list-style-type: none">• Boarded the ship together with the passenger• are related to the passenger either as spouse and/or sibling E.g. It is possible the passenger brought along his wife, brothers and sisters along a cruise ship |

| | | |
|----------------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | |
| NumParentChild | Number | <p>Whole number, the sum of the individuals who :</p> <ul style="list-style-type: none">• Boarded the ship together with the passenger• are related to the passenger either as parent and/or child <p>E.g. It is possible the passenger brought along his parents, and also his children along on the passenger ship</p> |
| Survived | Text | <p>“Yes” == passenger survived the ship accident</p> <p>“No” == passenger died</p> |
| | | |