Applied Machine Learning

Dr. Nassim Sohaee

Description

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

In this assignment, we ask you to complete the analysis of what sorts of people were likely to survive. In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.

Goal

It is your job to predict if a passenger survived the sinking of the Titanic or not. For each PassengerId in the test set, you must predict a 0 or 1 value for the Survived variable.

Metric

Your score is the percentage of passengers you correctly predict. This is known simply as "accuracy".

Submission File Format

You should submit a csv file with exactly 418 entries plus a header row. Your submission will show an error if you have extra columns (beyond PassengerId and Survived) or rows.

The file should have exactly 2 columns:

PassengerId (sorted in any order)

Survived (contains your binary predictions: 1 for survived, 0 for deceased)

```
PassengerId,Survived
 892,0
 893,1
 894,0
 Etc.
```

What you need to submit

1. A comprehensive report that shows what parameter you choose and compare your results to justify your choice of input parameters. Use graph and illustration to elaborate.
2. Python notebook.
3. Submission file described above.
4. Make sure to use similar graphs and plots we created in our python notebook in class to visualize data and predictive categories.

Project rules:

1. You can complete this project as a team project
2. Each team can have up to 2 members.
3. At the end of your report please specify the effort of each team member in completing this project. Here is a sample:

Team member 1	40%
Team member 2	60%
total	100%

Grading Rubric

Write-up

- Does the write-up explain the problem trying to solve? - 1 point
- Does the write-up explain the basics of machine learning algorithm used to solve the mentioned problem? - 1 point
- Does the write-up explain how specifically the machine learning algorithm was tuned/implemented? - 1 point
- Does the write-up has civilization component created in the python code? – 1 point
- Does the write-up has the comparison table to explain why the method of choice is better than other methods? – 1 point

Code

- Is the code in the working condition? – 1 point
- Is all the preprocessing and post processing is done in Python? – 1 point
- Is there a visualization component? – 1 point

Other:

- Output file format – 1 point
- Effort table – 1 point