Machine Learning

Project

One Titanic Problem

You work as an analyst for an insurance company that sells life insurance.

Your company has decided to specialize in cruise life insurance.

To avoid catastrophic financial losses, you have been tasked with determining how insurable is based on the likelihood of them surviving a cruise.

What better dataset to determine survivorship than one of the worst passenger ship disasters in history?

This isn’t just Big Data… It’s Titanic Data!

Enter the Titanic Survivors dataset: https://www.kaggle.com/c/titanic

Requirements:

Predict whether or not a passenger survived the Titanic

Work in groups of 2-3

Use at least 2 algorithms, show your work using a jupyter notebook

Generate a confusion matrix of your results. At minimum, calculate accuracy, recall, **F1-score**

Generate 2-3 visuals exploring your data and your results using Tableau

Present findings ~5 minutes

But Wait, There's More!

In true Kaggle fashion, we are offering a prize for the group with the highest F1-score!

Bragging rights!

(also a $10 dollar gift card to Starbucks for each member of the group)

Schedule:

November 17th (today):

-Intro

-Find Groups

-Working Session

November 26th/27th:

-First Half - Working Session

-Second Half - Presentations

Talking through the process:

What kind of question is being asked?

What kind of data do you have?

What are some possible algorithms you can use?

Additional Resources

Evaluation Metrics:

https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/

Feature Engineering Example:

https://triangleinequality.wordpress.com/2013/09/08/basic-feature-engineering-with-the-titanic-data/