SOCIETY OF ACTUARIES

2020 SOA Predictive Analytics 4.0

Hack-A-Thon Overview and Instructions

Predictive Analytics and Futurism Section
Bryon Robidoux FSA, CERA, MAAA

Limitations

The views expressed in this presentation are those of the presenters, and not those of the SOA. Nothing in this presentation is intended to represent a professional opinion or be an interpretation of actuarial standards of practice.

SOCIETY OF ACTUARIES

Goals of the Hack-A-Thon

- Get hands-on experience with predictive analytics for attendees less familiar with predictive analytics
- For those with more experience under their belt, provide a chance to test their skills in a friendly competition against peers
  - **Plus a chance to a badge of honor to be displayed in your email!**
- We can learn from the approaches and models that others are using
  - **Winners should plan to briefly discuss their methodology at the conclusion of the competition and coordinate with the SOA on a brief article for the next newsletter**
- For official rules, please refer to https://www.soa.org/link/f48f0b5775d14e1fb80c75059e50a302.aspx

SOCIETY OF ACTUARIES

Structure of the Competition

- A common training and testing data set will be provided to all teams
  - **Training set** includes response values
  - Response values for **testing set** are not available to teams
- Teams will be tasked with predicting the response variable for all observations in the testing set
  - We will provide a CSV file template in which you will submit your predictions
- A specific error metric will be calculated for each team's predictions
  - Best performance on the target error metric will determine the winner
  - Your method for arriving at the predictions is not considered for selecting the winner (no style points, sorry)
- After each team is able to successfully download the data, you will have **4 hours** to submit your predictions
- If a team submits multiple sets of predictions, the ones received latest will be used

What data will we be using?

- SOA student performance data
  - Student achievement in ASA exams
  - **16 data attributes** include demographic, social and behavioral features
  - Data is collected using reports and questionnaires
- Training set includes 589 instances
- Testing set includes 65 instances
- See data dictionary for more information

What are we predicting?

For each record in testing set,
will the student "pass" or "fail"?

- This can be identified using the target variable "**result**"
  - 1 = "pass", a student receives a pass when scoring 6 or greater
  - 0 = "fail", a student receives a fail when scoring 5 or lower

How will we be scored?

F1 Score = 2 x (precision * recall) / (precision + recall)
team with highest value wins

- Each prediction should be in the form of a factor ("pass" or "fail")
- The **F1 score** is between 0 and 1 and is the harmonic mean of precision and recall
- **Precision** answers the question: what proportion of predicted Positives (i.e. "pass") is truly Positive?
  - Precision = TP / (TP + FP)
- **Recall** answers the question: what proportion of actual Positives is correctly classified?
  - Recall = TP / (TP + FN)

	Actual Positive (Pass)	**Actual Negative (Fail)**
Predicted Positive (Pass)	True Positive (TP)	False Positive (FP)
Predicted Negative (Fail)	False Negative (FN)	True Negative (TN)

How do I submit my predictions?

- Enter prediction in template
  - Template can be downloaded via an electronic form; link shown below
  - Please ensure that your predictions align with the person IDs listed in the template – we will not be able to sort your data for you!
- Submit the template through the form available at https://soa.wufoo.com/forms/zl1f9vu18tvgdu/

SOCIETY OF ACTUARIES

OK, let's get going – how do I download the data?

- Data can be downloaded through a second electronic form:
  - https://soa.wufoo.com/forms/2020-hackathon-files/

Once all teams have successfully downloaded the data, we will start the clock!

- The original data is called "Student Performance Data Set" and it's taken from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/student+performance

SOCIETY OF ACTUARIES

Questions?

- **Thank you and have fun!**

SOCIETY OF ACTUARIES