

Marathon Match

[Problem Statement](#)Contest: [Fishing MM](#)[Register & Rules](#) | [Standings/Registrants](#) | [Forum](#) | [Printable view](#)Problem: [FishingForFishermen](#)

Problem Statement

Prize Distribution

- 1st - \$9000
- 2nd - \$4000
- 3rd - \$2500
- 4th - \$1500
- 5th - \$1000
- Total - \$18000

Bonus Prize

We will ask the 5 winners to submit a [Docker](#) file for their solution, and, if so will provide guidance in that regard. We will pay the winners who submit a Docker file an additional \$100 each for their efforts.

Requirements to Win a Prize

In order to receive a prize, you must do all the following:

1. Achieve a score in the top 5, according to system test results. See the "Scoring" section below.
2. Create a legitimate algorithm that runs successfully on a different data set with the same fields. Hard-coded solutions are unacceptable.
3. Within 7 days from the end of the challenge, submit a complete report at least 2 pages long outlining your final algorithm, explaining the logic behind and steps to its approach. The required content and format appear in the "Report" section below.
4. Within 7 days of the end of the challenge, submit all code used in your final algorithm in 1 appropriately named file (or tar or zip archive). We will contact the winners via email and ask for the file. The naming convention should be memberHandle-FishingForFishermenContest. For example, handle "johndoe" would name his submission "johndoe-FishingForFishermenContest."

If you place in the top 5 but fail to do any of the above, then the corresponding prize money will be added to the prize pool of the next Fishing for Fishermen contest.

Background

NASA wishes to identify whether a vessel is fishing or not based on AIS broadcast reports and contextual data.

Objective

Create an algorithm to effectively identify if a vessel is fishing based on observable behavior and any additional data such as weather, known fishing grounds, etc. regardless of vessel type or declared purpose.

Your algorithm shall utilize:

- AIS positional data
- Other data from openly accessible data sources that can be used to determine fishing behavior

Your algorithm should then detect vessels that match the profile of behaviors of vessels engaged in fishing. You must identify whether each vessel is fishing or not at each point in time specified in the data set.

Data Description

The data set is provided in as a series of AIS messages in NMEA format, each message having one appended field indicating, in Unix epoch time format, the "time of fix" (the time at which each AIS broadcast was received). Some messages may contain duplicates or almost duplicates in the data.

Example AIS messages with the appended Time of Fix:

```

"!AIVDM,1,1,,A,15NVOL?002o5KFRJ3Rh@0:;p0000,0*73",1439479221
"!AIVDM,1,1,,A,1EOP=17018o61GPJ>owGtFD>0>@<,0*21",1439698812
"!AIVDM,1,1,,B,BE2P9a@15eib9RVWI:4iWws5oP00,0*2F",1438900139
"!AIVDM,1,1,,A,35DCJ1100QoLlQ8BIKN<P:0D0>@<,0*19",1438478590
"!AIVDM,1,1,,A,15MC0:000019La1OLF:bk8bH0861,0*7B",1437222010
"!AIVDM,1,1,,B,169EG;@P1:nTn1Gm25U`wwJ28KB,0*6A",1435876966
"!AIVDM,1,1,,A,14W9Ip0017GS>8dB<KvbA8N00<1=,0*22",1439466483
"!AIVDM,1,1,,B,15N4r6P01FD2R>6O?n;F54WR0D04,0*33",1435748808

```

```
"!AIVDM,1,1,,A,15RSMr001p9ho>B4CSBM3:H0h0P,0*4D",1441389209
"!AIVDM,1,1,,B,1HL20h001qo2Uu6JQi;0HkN0>`<,0*5E",1439727293
```

Some messages may have been garbled or corrupted during transmission, resulting in incorrect checksums or message lengths. A link in the resources section is provided to calculate the checksum of each message.

The data is split into two files: one for training available [here](#) and another for testing available [here](#). The provided training data file contains the raw AIS position report messages along with the decoded information. The ground truth for each of these rows is in the final column (FISHING_STATUS). The training data file has the following columns:

- **TIMESTAMP** - Time in Unix epoch format
- **TIME** - Time in human readable format
- **MESSAGE_ID** - Message type
- **MMSI** - Vessel identification number (not necessarily unique to a s
- **LAT** - Current latitude
- **LONG** - Current longitude
- **POS_ACCURACY** - Indicates position accuracy
- **NAV_STATUS** - Indicates navigational status
- **SOG** - Speed over ground
- **COG** - Course over ground
- **TRUE_HEADING** - Heading
- **RAW_MESSAGE** - Full raw AIS position report message
- **FISHING_STATUS** - Fishing status (Fishing, Not Fishing, or Unknown)

The provided testing data file will only contain the columns **TIMESTAMP** and **RAW_MESSAGE**. Your program should read in data from the provided data files, and output a csv file containing the confidence predictions of the fishing status for each vessel at each point in time, one for every row in the testing data set. For example, a fishing confidence of 0.0 indicates that the vessel is definitely not fishing at this point in time while a fishing confidence of 1.0 indicates that the vessel is definitely fishing at this point in time. Your output file should contain only one confidence prediction per line in the same order as the provided testing data file. Line N of your output file corresponds to the Nth message (line) in the provided testing data file. Each line should contain only one human readable number between 0 and 1 indicating the fishing confidence for the corresponding message.

An additional file containing AIS static report messages can be downloaded [here](#). You may be able to use the additional vessel information in these messages to improve your predictions. This file may not contain information on all vessels in the training and testing files. This file is not used for scoring.

Functions

During the contest, only your results will be submitted. You will submit code which implements only one function, **getAnswerURL()**. Your function will return a String corresponding to the URL of your answer .csv file. You may upload your .csv file to a cloud hosting service (such as [Dropbox](#)) which can provide a direct link to your .csv file. To create a direct sharing link in Dropbox, right click on the uploaded file and select share. You should be able to copy a link to this specific file which ends with the tag "?dl=0". This URL will point directly to your file if you change this tag to "?dl=1". You can then use this link in your **getAnswerURL()** function. Your complete code that generates these results will be tested at the end of the contest before prize distribution.

Scoring

Your predictions will be scored against the ground truth using the area under the [receiver operating characteristic](#) (ROC). You must return a fishing confidence prediction for every candidate in the test data set in order to receive any points.

The ROC curve will be determined and the score will be determined from the area under the ROC curve using the following method:

1. The contestant will score each AIS message with a confidence that the vessel is fishing at the time the message was broadcast. These will constitute the CSV file described above, having one field: Confidence_j, where j=1, ..., M. Independently, Topcoder will add a second column, GroundTruth_j, where GroundTruth_j=1 for "Fishing" and 0 for "Not Fishing", based on all-source analysis by the U.S. Coast Guard Fisheries analysts. Lines where GroundTruth is "Unknown" will be eliminated, resulting in a modified CSV file with N lines (N≤M), two fields in each line.
 2. The modified CSV file will be sorted by the contestant's confidence field from largest to smallest. In case of ties in confidence, tied rows will be kept in their original order (the order of the testing data file). This results in a new sequence of GroundTruth values, s_i (where i=1, ..., N).
2. The true positive rates and the false positive rates are determined as follows:

$$TPR_i = \text{Accumulate}[s_i] / N_TPR$$

$$FPR_i = \text{Accumulate}[1 - s_i] / N_FPR;$$

with the addition: **FPR₀** = 0;

where N_{TRP} is the total number of fishing records, N_{FPR} is the total number of non-fishing records, and $N_{TRP} + N_{FPR} = N$ (total number of records with known status in the test)

3. Then the **AuC** is determined as a numerical integral of **TRP** over **FRP**:

$$AuC = \text{Sum} [TPR_i * (FPR_i - FPR_{i-1})]$$

4. Then the **AuC** is scaled to determine the final score:

$$\text{Score} = \max(1,000,000 * (2 * AuC - 1), 0)$$

Resources

Useful site for the entire AIS message (each line of data to be provided is an AIS "sentence"):

http://catb.org/gpsd/AIVDM.html#_aivdm_aivdo_sentence_layer

ITU document describing "payload" field (field 6 of each message):<https://www.itu.int/rec/R-REC-M.1371/en>

How to calculate the NMEA checksum: <https://rietman.wordpress.com/2008/09/25/how-to-calculate-the-nmea-checksum/>

The distance between 2 points, given their longitude and latitude, can be calculated using the [Haversine formula](#).

Report

Your report must be at least 2 pages long, contain at least the following sections, and use the section and bullet names below.

Your Information

This section must contain at least the following:

- First name
- Last name
- Topcoder handle
- Email address

Approach Used

Please describe your algorithm so that we know what you did even before seeing your code. Use line references to refer to specific portions of your code.

This section must contain at least the following:

- Approaches considered
- Approach ultimately chosen
- Steps to approach ultimately chosen, including references to specific lines of your code
- Open source resources and tools used, including URLs and references to specific lines of your code
- Advantages and disadvantages of the approach chosen
- Comments on libraries
- Comments on open source resources used
- Special guidance given to algorithm based on training
- Potential improvements to your algorithm

Definition

Class: FishingForFishermen
 Method: getAnswerURL
 Parameters:
 Returns: string
 Method signature: def getAnswerURL(self,): (be sure your method is public)

Examples

```
0)
Seed: 0
1)
Seed: 0
2)
```

Seed: 0

This problem statement is the exclusive and proprietary property of TopCoder, Inc. Any unauthorized use or reproduction of this information without the prior written consent of TopCoder, Inc. is strictly prohibited. (c)2010, TopCoder, Inc. All rights reserved.

[OTHERS](#)

[SITEMAP](#)

[ABOUT US](#)

[CONTACT US](#)

[HELP CENTER](#)

[PRIVACY POLICY](#)

[TERMS](#)

topcoder is also on

© 2015 topcoder. All Rights Reserved