



# R&D ML Workshop SMS Classification

ROCS | Kayla Dubnick | dubnickk@vmware.com

2018-08-03 - 2018-08-31

The training set is 5,074 SMS messages that are either ham or spam.

In **sms-train.csv**:

Column 1 is the SMS message.

Column 2 is the label. It's either "Ham" or "Spam".

Your task is to build a classifier the correctly classifies the SMS messages to match the labels.

In **sms-eval.csv**:

Column 1 is the ID of the evaluation SMS.

Column 2 is the text of the evaluation SMS.

**Output evaluation file format:**

Column 1 is the ID from sms-eval.csv.

Column 2 is your predicted class, either "Ham", or "Spam".

**Output Example:**

```
ID, HAM-SPAM
1, Ham
2, Ham
3, Spam
4, Ham
...
```

Good luck!

## Leaderboard

#	Name	Kernel	Score	Entries
1	Alex RN	TFIDF	0.87000	1
2	Rick Battle	Coin Flip	0.54800	1

[SUBMIT ANSWERS](#)[SEE ALL ENTRIES](#)

# Files

[classify.py](#) (3.4 KB)

[eval.py](#) (1.2 KB)

[sms-eval.csv](#) (54.2 KB)

[sms-train.csv](#) (415.3 KB)

# Comments

What are your thoughts?

ADD COMMENT