



<https://github.com/zshamrock/nlp-getting-started>

# Real or Not?

# Evaluate

Submissions are evaluated using [F1](#) between the predicted and expected answers.

F1 is calculated as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

where:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

and:

True Positive [TP] = your prediction is 1, and the ground truth is also 1 - you predicted a *positive* and that's *true*!

False Positive [FP] = your prediction is 1, and the ground truth is 0 - you predicted a *positive*, and that's *false*.

False Negative [FN] = your prediction is 0, and the ground truth is 1 - you predicted a *negative*, and that's *false*.

## Submission File


For each ID in the test set, you must predict 1 if the tweet is describing a real disaster, and 0 otherwise. The file should contain a header and have the following format:

```
id,target
0,0
2,0
3,1
9,0
11,0
```

# Scores

Submission and Description	Public Score	Use for Final Score
<a href="#">03_google_automl.csv</a> 2 minutes ago by Aliaksandr Kazlou 03 Google AutoML	0.81186	<input type="checkbox"/>
<a href="#">04_lstm_submission.csv</a> an hour ago by Aliaksandr Kazlou 04 Simple LSTM	0.76380	<input type="checkbox"/>
<a href="#">02_fasttext_submission.csv</a> 7 hours ago by Aliaksandr Kazlou 02 FastText supervised, no hyperparameters search	0.74437	<input type="checkbox"/>
<a href="#">02_fasttext_submission.csv</a> 7 hours ago by Aliaksandr Kazlou 02 FastText <a href="https://fasttext.cc/">https://fasttext.cc/</a>	0.52044	<input type="checkbox"/>
<a href="#">01_classic_text_classification_submission.csv</a> 9 hours ago by Aliaksandr Kazlou 01 Classic Text Classification (CountVectorizer + RidgeClassifier), default, no hyperparameters search, no features engineering	0.77096	<input type="checkbox"/>


# Leaderboard


1204	Aliaksandr Kazlou		0.81186	5	3m
------	-------------------	-------------------------------------------------------------------------------------	---------	---	----

Your Best Entry ↑

You advanced 1,604 places on the leaderboard!

Your submission scored 0.81186, which is an improvement of your previous score of 0.77096. Great job!

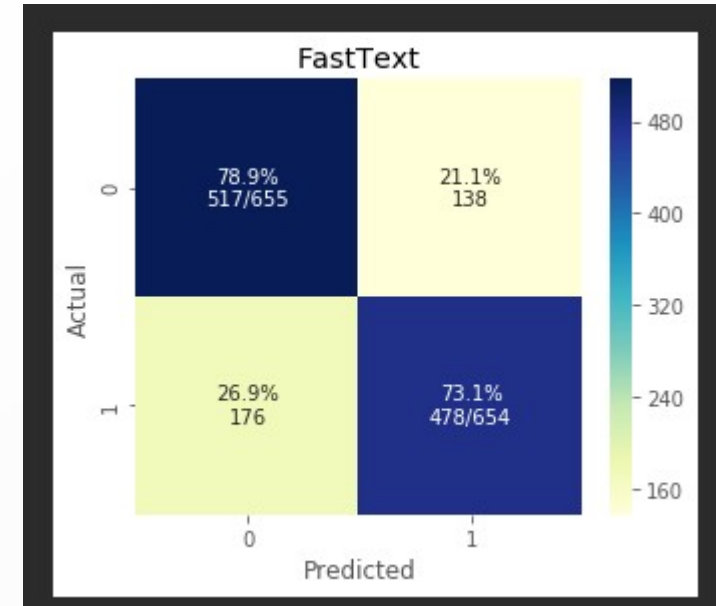
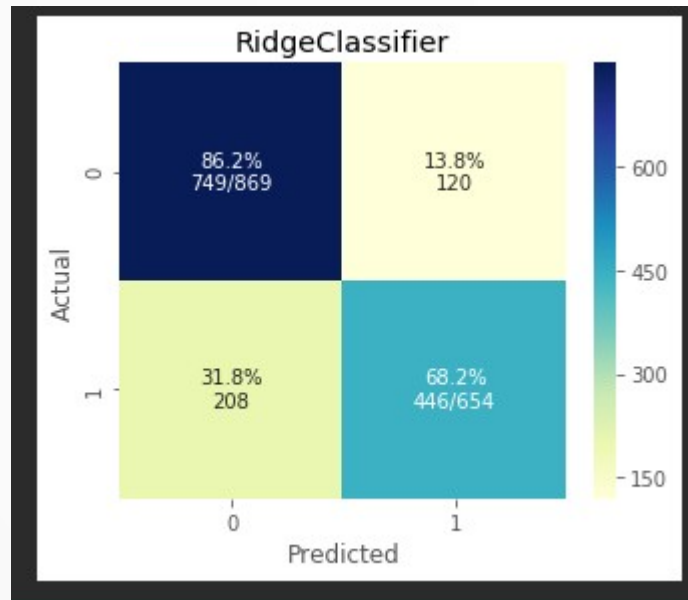
 Tweet this!

	AutoML Benchmark	0.81083
-----------------------------------------------------------------------------------	------------------	---------

# Models

- Linear models – RidgeClassifier (0.77096<sup>2</sup>)
- FastText (0.74437<sup>4</sup>)
- Google AutoML (0.81186<sup>1</sup>)
- LSTM (0.76380<sup>3</sup>)

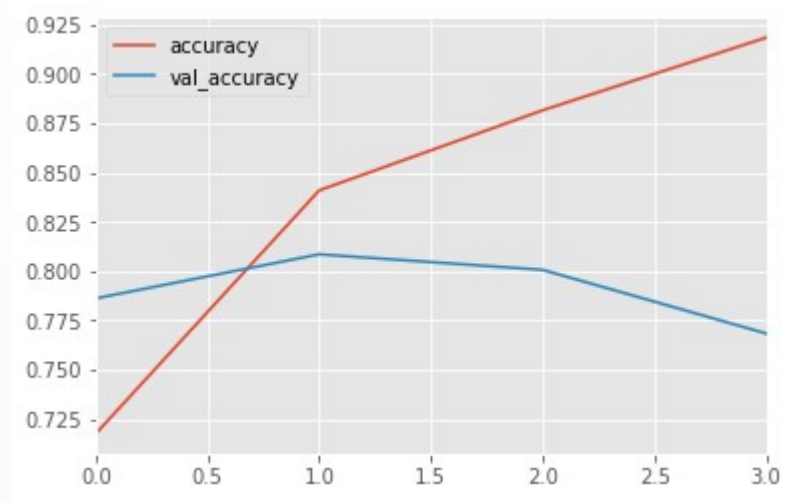
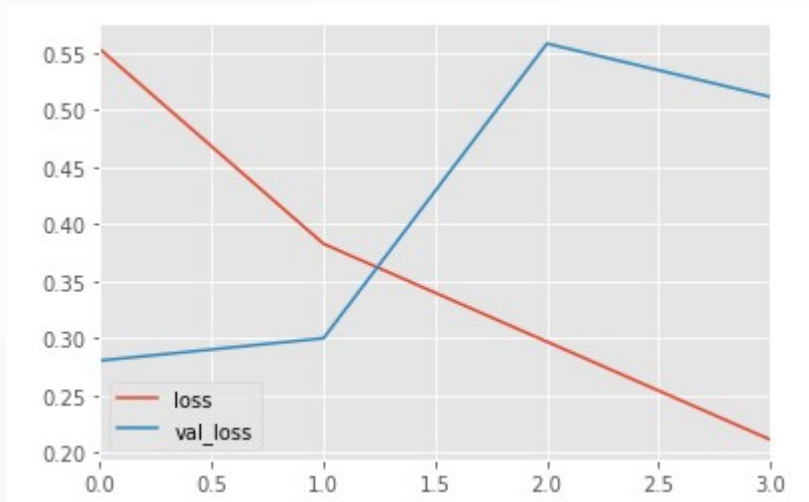
# Confusion Matrix






True label	Predicted label	
	0	1
0	88%	12%
1	30%	70%



# LSTM Evaluate

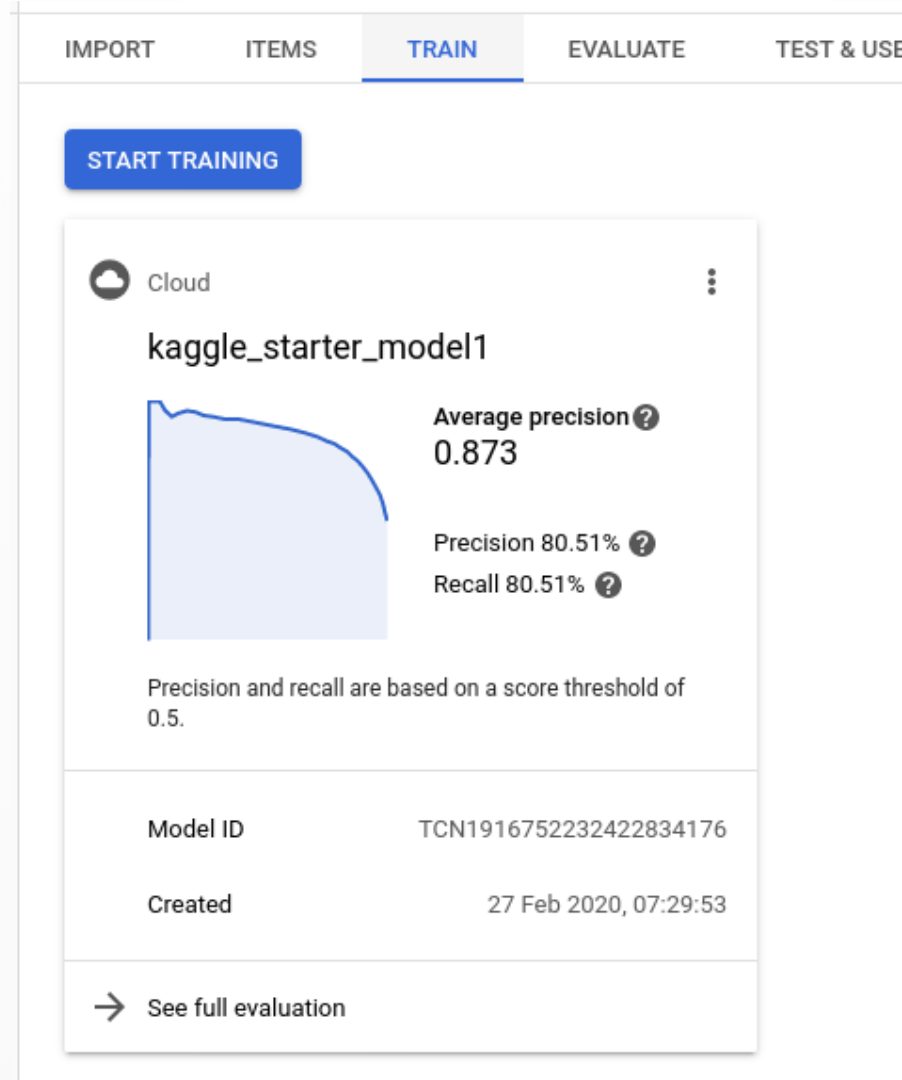


# Google AutoML: items

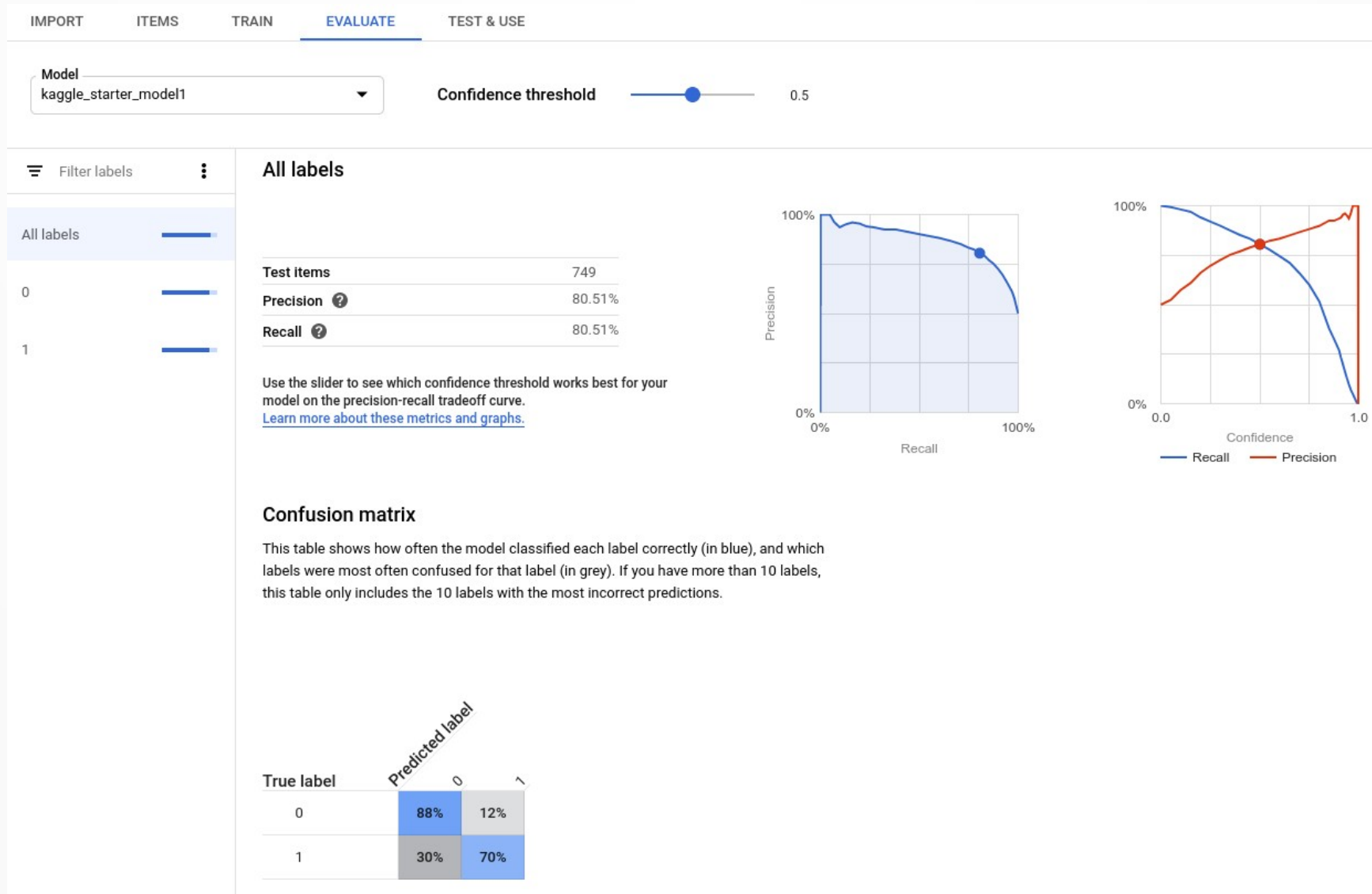
IMPORT	ITEMS	TRAIN	EVALUATE	TEST & USE	Single-label classification
All items	7,503	<div>  Warning: Importing data           <a href="#">DETAILS</a> <a href="#">DISMISS</a> </div>			
Labelled	7,485	<div>  Filter table            </div>			
Unlabelled	18	<div> <input type="checkbox"/> Items           Labels         </div>			
Training	5,987	<div> <input type="checkbox"/> 70 years since we annihilated 100000 people instantly and became aware that we have the ability to a...           1         </div>			
Validation	749	<div> <input type="checkbox"/> Stu Dorret's mudslide rubber tyre cake may have saved you #GBBO           0         </div>			
Testing	749	<div> <input type="checkbox"/> her cake looks like a mudslide hah           0         </div>			
		<div> <input type="checkbox"/> http://t.co/kg5pLkeDhr WRAPUP 2-U.S. cable TV companies' shares crushed after Disney disappoints htt...           0         </div>			
		<div> <input type="checkbox"/> Reddit Will Now Quarantine Offensive Content: Reddit co-founder and CEO Steve Huffman has unveiled m...           0         </div>			
		<div> <input type="checkbox"/> This desperation dislocation Separation condemnation Revelation in temptation Isolation desolation L...           1         </div>			
0	4,297	<div> <input type="checkbox"/> Homecoming Queen Killed on Way Home from the Prom by Flood Waters! #socialnews http://t.co/VmKexjTyG...           1         </div>			
1	3,188	<div> <input type="checkbox"/> Malik Saadthe best police officer produced by KPdied in a suicide bombing on 27 Jan 2007he was more...           1         </div>			
		<div> <input type="checkbox"/> Guys. I have an Imouto Who Isn't Actually Related to Me by Blood.           0         </div>			
		<div> <input type="checkbox"/> @cityofhummus @ILNewsFlash do you want to hear more? Displaced my parents until both died in the dia...           0         </div>			
ADD NEW LABEL					



# Google AutoML: train



# Google AutoML: evaluate



# Next Steps

- General
  - Features engineering
  - Preprocessing
- Linear Models
  - Different classification models, ensembles (random forest, gradient boosting), stacking/voting
- FastText
  - Following official latest supervised tutorial
  - Automatic hyperparameters optimization

# Next Steps (contd.)

- LSTM
  - More complex LSTM with Embeddings
- Other Models
  - GloVe
  - BERT
  - H2O AutoML