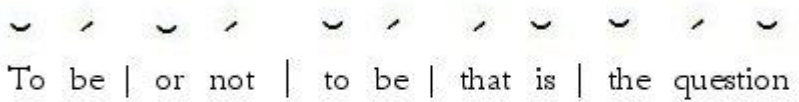
Teaching Machines to Identify Sonnets

By Michael A. Schulze

What's Iambic Pentameter?

- 10 - 12 syllables
- 5 meters or foots
- Basic Iambic: One short (unstressed) syllable followed by one long (or stressed) syllable.
- **But** there are many exceptions to the rule...

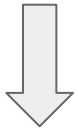


Exceptions and Difficulties

- Feminine ending: 11 syllables in a line with the last extra syllable being unstressed
- Inversion: Switches the order of the stresses so the stress is on the second syllable instead of the first.
- Caesura: Two stressed syllables in the same iamb to break up a line or a thought.
- Falling meter: The first iamb stresses are switched so the first syllable is stressed and the second unstressed, but unlike inversion this only lasts for the first iamb.
- Different meters for the same word.

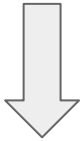
Break it Down

```
['But', 'as', 'the', 'riper', 'should', 'by', 'time', 'decease']
```

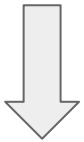


```
1 But [u'B', u'AH1', u'T'] 1
1 as [u'AE1', u'Z'] 1
1 as [u'EH1', u'Z'] 1
0 the [u'DH', u'AH0'] 1
1 the [u'DH', u'AH1'] 1
0 the [u'DH', u'IY0'] 1
riper ---> NOT IN CMU DICT <---
1 should [u'SH', u'UH1', u'D'] 1
1 by [u'B', u'AY1'] 1
1 time [u'T', u'AY1', u'M'] 1
0 decease [u'D', u'IH0', u'S', u'IY1', u'S'] 1
1 decease [u'D', u'IH0', u'S', u'IY1', u'S'] 2
```

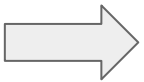
Transforming...



```
[['But', [['But0', [1]]]],
 ['as', [['as0', [1]], ['as1', [1]]]],
 ['the', [['the0', [0]], ['the1', [1]], ['the2', [0]]]],
 ['riper', [['riper0', [-1, -1]]]],
 ['should', [['should0', [1]]]],
 ['by', [['by0', [1]]]],
 ['time', [['time0', [1]]]],
 ['decease', [['decease0', [0, 1]]]]
```



```
[['But', ['But', [1]]],
 ['as', ['as', [1]]],
 ['the', ['the', [0]]],
 ['riper', ['riper', [-1, -1]]],
 ['should', ['should', [1]]],
 ['by', ['by', [1]]],
 ['time', ['time', [1]]],
 ['decease', ['decease', [0, 1]]]]
```



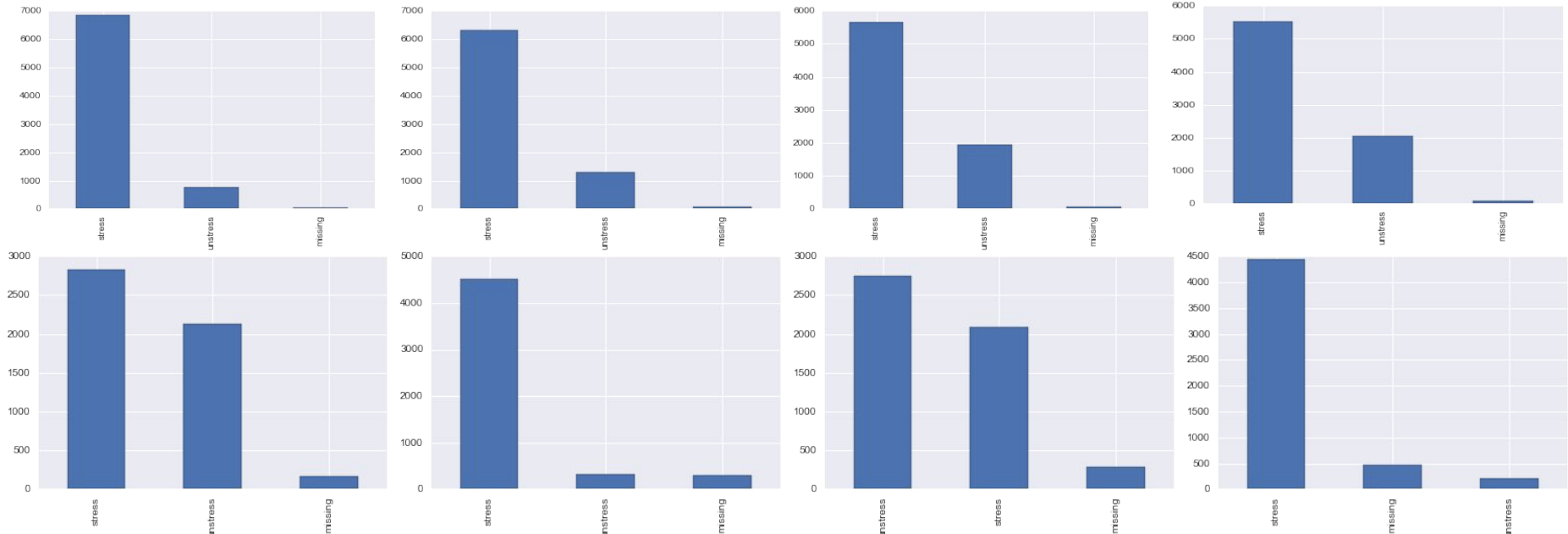
```
[1, 1, 1, 2, 1, 1, 1, 2]
```

Transformed!

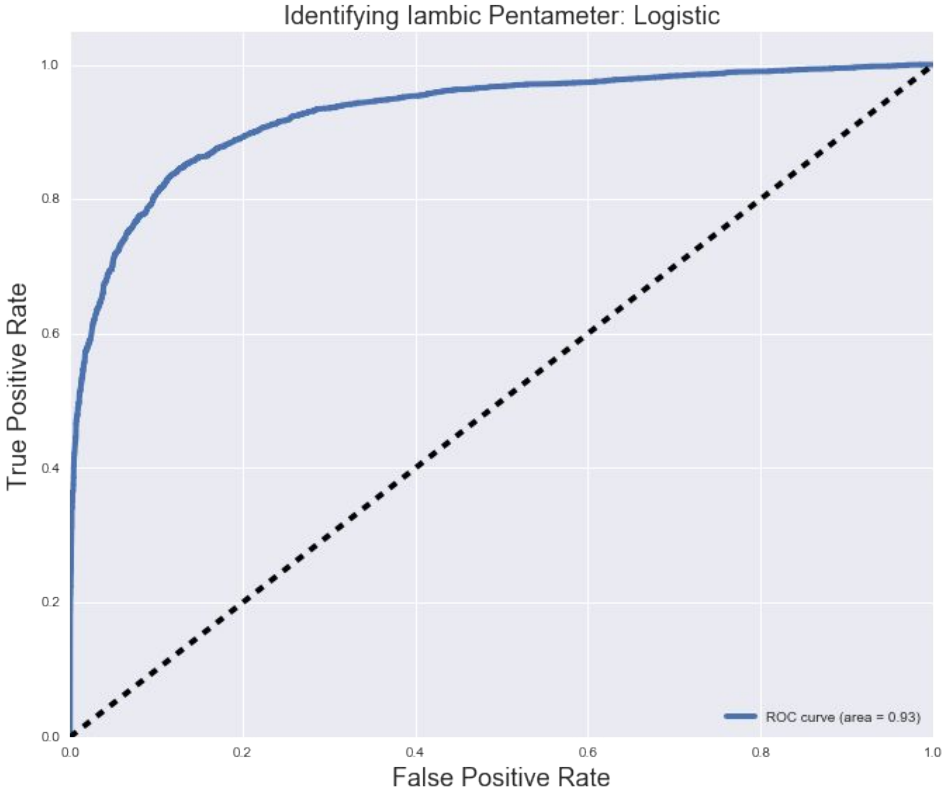
```
[['From', ['From', [0]]], ['fairest', ['fairest', [1, 0]]], ['creatures', ['creatures', [1, 0]]], ['we', ['we',
 [1]]], ['desire', ['desire', [0, 1]]], ['increase', ['increase', [0, 1]]]]

[[0], [1, 0], [1, 0], [1], [0, 1], [0, 1]]

[0, 1, 0, 1, 0, 1, 0, 1, 0, 1]
```



Performance

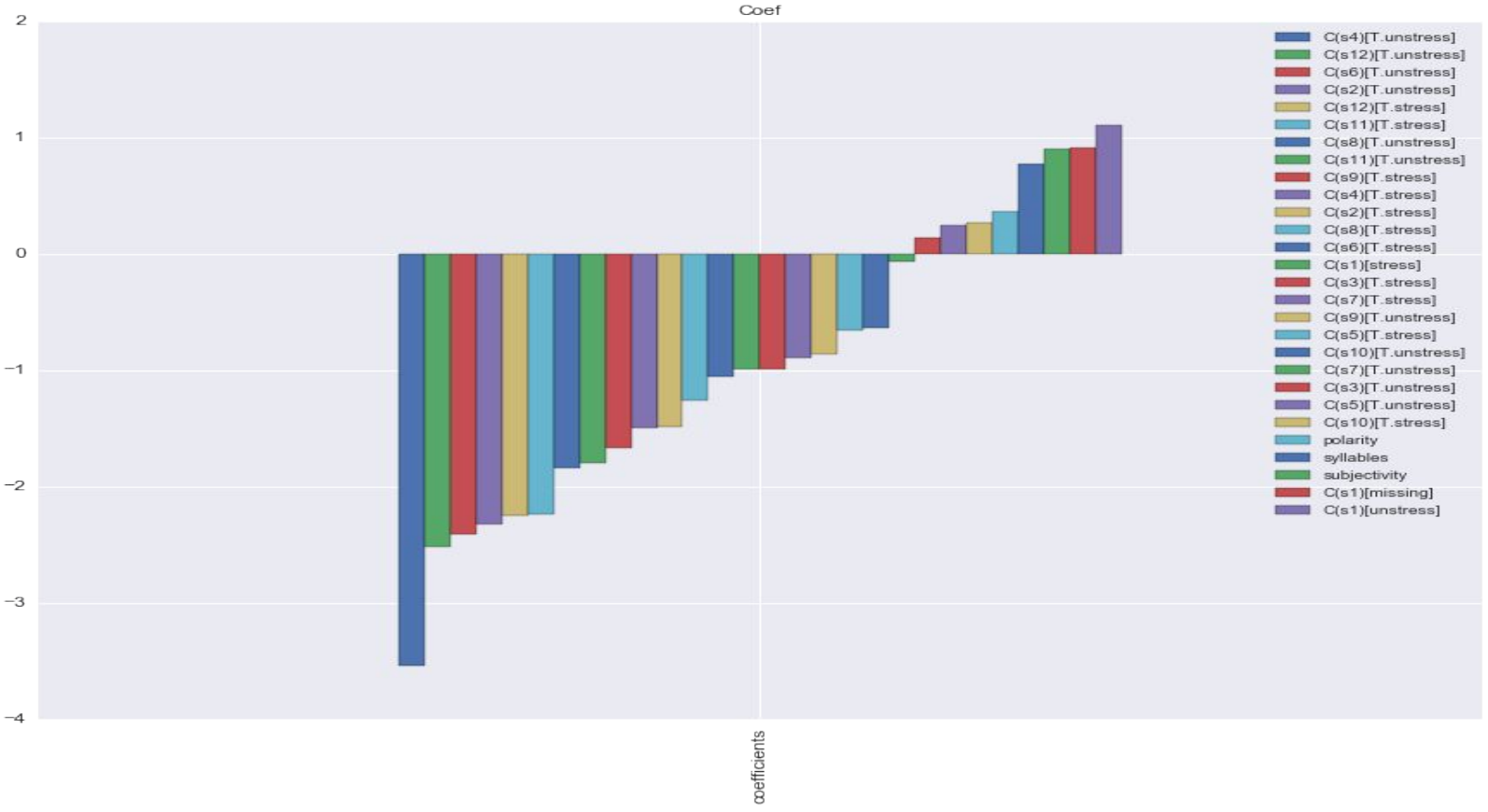


Predicted	0.0	1.0	All
True			
0.0	6421	762	7183
1.0	937	4182	5119
All	7358	4944	12302

0.893916191007

0.816956436804

Explanatory



The Future

- Improve predictability through more data and identifying cases where the text is actually iambic and adding those cases back into my training set.

- Using the model to answer if writing in iambic (subconsciously or consciously) results in more popular lyrics, tweets, or news titles and if certain people use iambic more or less often than others.

Question?