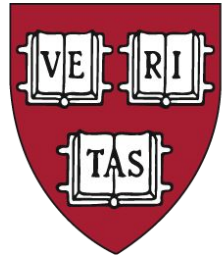

Enhanced AI Scoring Engine

Investigation by Kenny Leung
October 17, 2017

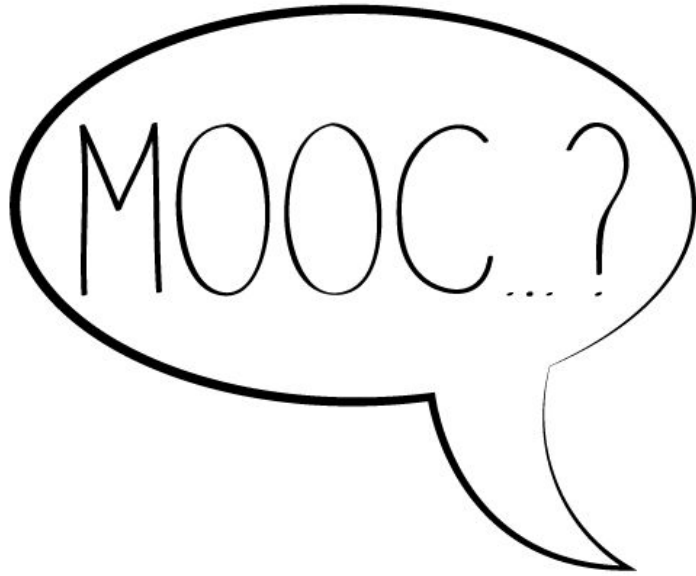


**Massachusetts
Institute of
Technology**

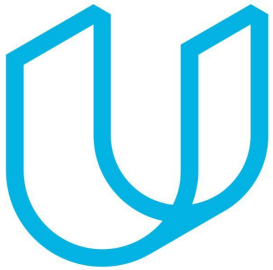


HARVARD
UNIVERSITY

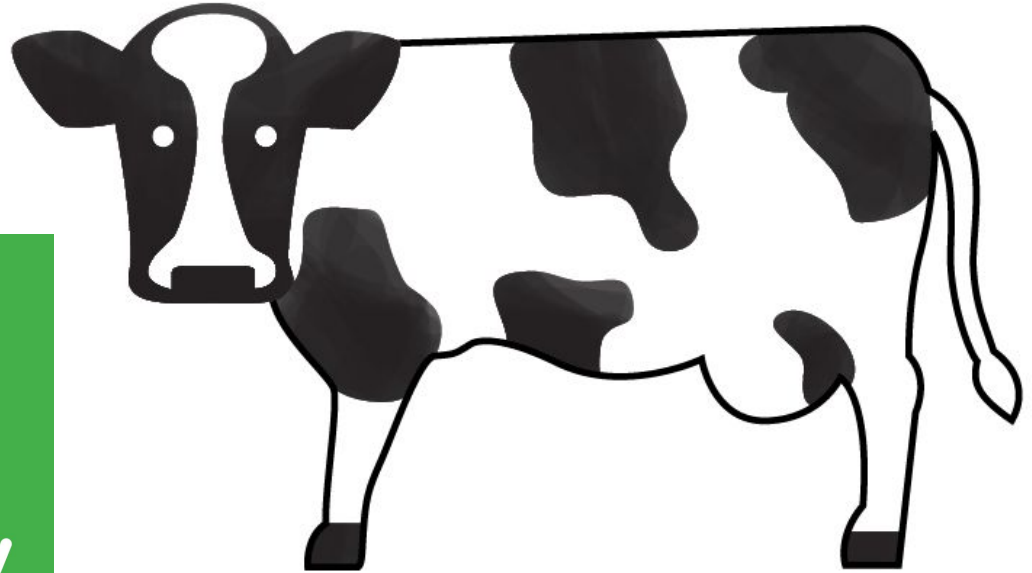




coursera



UDACITY



edx / ease

Watch 165 Star 127 Fork 68

Code Issues 1 Pull requests 1 Projects 0 Wiki Insights

Branch: master ease / ease /

Create new file Upload files Find file History

wedaly Use a set instead of a list for good ngram lookup

Latest commit 94014b7 on Jun 23, 2014

..		
data	Remove NLTK corpus from repository; download it from S3 instead	3 years ago
external_code	Remove fisher, fix generic model generation, update tests	4 years ago
tests	Added tests for spellcheck	4 years ago
__init__.py	Rename, strip out statsd	5 years ago
create.py	Add in test case creation, ability to run those test cases	4 years ago
essay_set.py	Remove random.seed call in essay_set	3 years ago
feature_extractor.py	Use a set instead of a list for good ngram lookup	3 years ago
grade.py	Add in test case creation, ability to run those test cases	4 years ago
model_creator.py	Change error logging for an error we're suppressing to debug logging	3 years ago
predictor_extractor.py	Rename, strip out statsd	5 years ago
predictor_set.py	Remove fisher, fix generic model generation, update tests	4 years ago
util_functions.py	Create unique tempfiles in the system /tmp directory	4 years ago

Enhanced

AI

Scoring

Engine

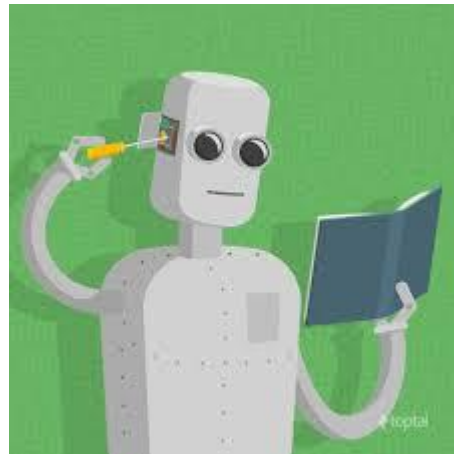
Input Data

- Input essays with integer scores
- Strip non-ASCII characters



Text Pre-processing

- Removes punctuation
- Truncates essays to 20,000 words
- Runs spell-check



Enhanced

AI

Scoring

Engine

Natural Language Toolkit (nltk)

Tokenize essay corpus

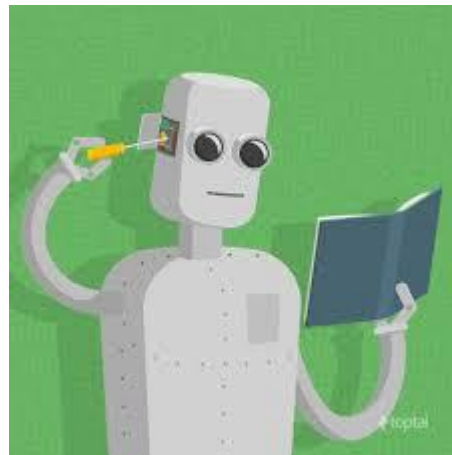
- `nltk.word_tokenize()`

Tag parts of speech

- `nltk.pos_tag()`

Stem the tokens

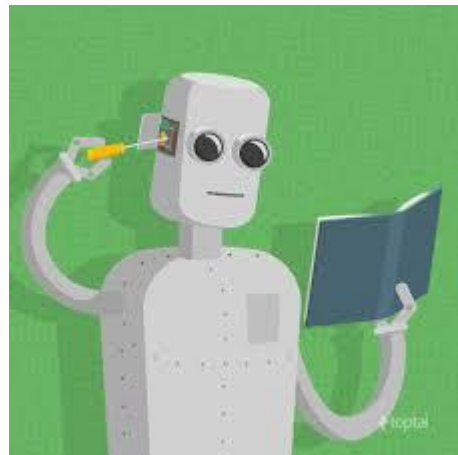
- `nltk.PorterStemmer()`
-



Enhanced AI Scoring Engine

SciKit Learn

1. Vectorization
2. Term weighting & Normalization
3. Add features
4. Model





Opponents of Automated Essay Scoring

Resources

<http://vikparuchuri.github.io/boston-python-ml/#1>

<http://mitros.org/p/papers/linc13.pdf>

Questions?
