

words = 47  
 total\_characters = 189  
 spaces = 42  
 different\_words = 22  
 avg\_word\_len = 2.98  
 median\_word\_len = 3  
 std\_dev\_word\_len = .94  
 avg\_sentence\_len = 15.67  
 median\_sentence\_len = 17  
 std\_dev\_sentence\_len = 9.07  
 lexical\_density = 95.7  
 readability = 6.3  
 avg\_syllables = 1.14  
 sentence\_count = 4  
 avg\_sentence\_length = 15.67  
 max\_sentence\_len = 24  
 min\_sentence\_len = 6  
 complex\_words = 0  
 one\_syllable\_word = 38  
 two\_syllable\_word = 6

polarity = 0.19  
 subjectivity = 0.38  
 the\_count = 5  
 to\_count = 5  
 and\_count = 5  
 modality = 0.75  
 noun\_chunks = 12  
 person = 1  
 cardinal = 1  
 quotation\_mark\_count = 0  
 apostrophe\_count = 0  
 colon\_count = 0  
 semicolon\_count = 0  
 question\_mark\_count = 0  
 exclamation\_mark\_count = 0  
 dash\_count = 0  
 parenthesis\_count = 0  
 bracket\_count = 0  
 brace\_count = 0

agreeableness = 0.14  
 analytical = 0.1  
 anger = 0.27  
 confident = 0  
 conscientiousness = 0.65  
 disgust = 0.05  
 emotional\_range = 0.79  
 extraversion = 0.02  
 fear = 0.12  
 joy = 0.1  
 openness = 0.17  
 sadness = 0.56  
 tentative = 0.66

part\_speech\_comma = 4  
 part\_speech\_period = 3  
 part\_speech\_cc = 3  
 part\_speech\_cd = 1  
 part\_speech\_dt = 7  
 part\_speech\_in = 4  
 part\_speech\_jj = 3  
 part\_speech\_nn = 7  
 part\_speech\_nnp = 2  
 part\_speech\_nns = 1  
 part\_speech\_prp = 3  
 part\_speech\_rb = 1  
 part\_speech\_to = 5  
 part\_speech\_uh = 1  
 part\_speech\_vb = 5  
 part\_speech\_vbd = 1  
 part\_speech\_vbn = 1  
 part\_speech\_vbp = 1  
 part\_speech\_vbz = 1

gunning\_fog = 6.27  
 flesch-kincaid = 3.07  
 smog = 3.13  
 coleman\_liau = 0.7  
 automated = 1.13  
 average\_grade\_level = 2.86  
 median\_grade\_level = 3.07  
 nouns = 0.234  
 adjectives = 0.0851  
 verbs = 0.1277  
 adverbs = 0.0213  
 prepositions = 0.0851  
 pronouns = 0.0638  
 auxiliary\_verbs = 0.0638  
 passive\_voice = 0

words\_per\_minute = 45  
 mistyped\_words = 1  
 corrected\_mistakes = 5  
 characters\_per\_minute = 230  
 fastest\_finger = left\_pinky  
 slowest\_finger = right\_middle  
 word\_accuracy = 99%  
 language = english  
 letter\_casing = mixed  
 type\_duration = 23\_seconds  
 left\_pinky = 171ms  
 left\_ring = 200ms  
 left\_middle = 220ms  
 left\_index = 177ms  
 right\_index = 245ms  
 right\_middle = 643ms  
 right\_ring = 383ms  
 right\_pinky = 467ms  
 thumbs = 218ms

time\_to\_mistake = 545ms  
 punctuation\_pause = 100ms  
 capital\_letter\_pause = 35ms  
 exclamation\_pause = null

many more...

many more...

Paragraph Mechanics

Typing Performance

Sentence #82

Dear Dan,  
 Yes, I came out of the corn, back to the city,  
 both to draw and to do copy on the new cars. To  
 date I am able to put cash in the bank and bear a  
 bill or two. The new deed has done it.

this is the hardest part of machine learning

sentence_number	words	total_characters	spaces	different_words	avg_word_len	median_word_len	std_dev_word_len	avg_sentence_len	median_sentence_len	std_dev_sentence_len	lexical_density	readability
82	47	189	42	22	2.98	3	0.94	15.67	17	9.07	95.7	6.3

Feature Extraction

Banding

model needs limited choices (up to 5) to make accurate predictions

primary key	outcome	some will be letters	some will be 1 to 5 columns	1 to 3	1 to 2	0 to 1								
sentence_number	label	level	words	total_characters	spaces	different_words	avg_word_len	median_word_len	std_dev_word_len	avg_sentence_len	median_sentence_len	std_dev_sentence_len	lexical_density	readability
82	0	A	2	2	3	1	1	1	1	3	1	2	2	2
110	1	C	2	5	1	2	3	1	1	2	2	2	1	0
98	1	A	4	3	2	2	2	1	1	1	3	2	1	0
66	1	A	4	2	2	2	2	1	1	1	2	2	0	0
4	0	D	1	3	3	2	5	0	1	1	2	2	1	1

... 100s of features

Model Setup

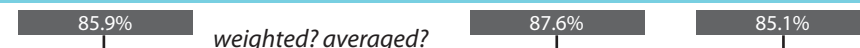
Machine Models



can evaluate existing sentences  
 score = 1, probability = 90%+  
 could be in early lesson

score = 1, probability = 60%  
 could be in advanced lesson

Scores



Web Service

new sentence	label	level	words	total_characters	spaces	different_words	avg_word_len	median_word_len	std_dev_word_len	avg_sentence_len	median_sentence_len	std_dev_sentence_len	lexical_density	readability	score	probability
X	0	A	2	2	3	1	1	1	1	3	1	2	2	2	1	0.865518

models can be built for different lesson levels

is this a false positive? no. single measure said bad sentence. 100+ features said, no, it's really a good sentence. trust the model

Adaptive Learning

note: new sentences can only be evaluated against mechanics, since performance does not exist