Automated Essay Grading

Using Neural Networks

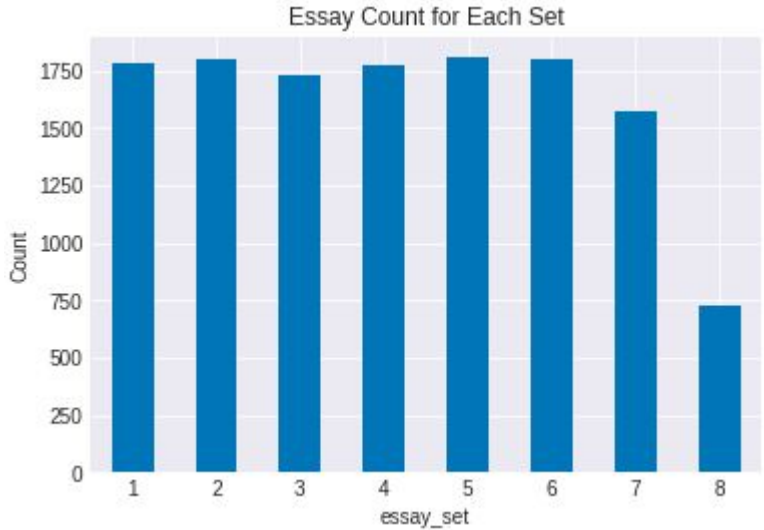
Revathi Bhuvaneswari

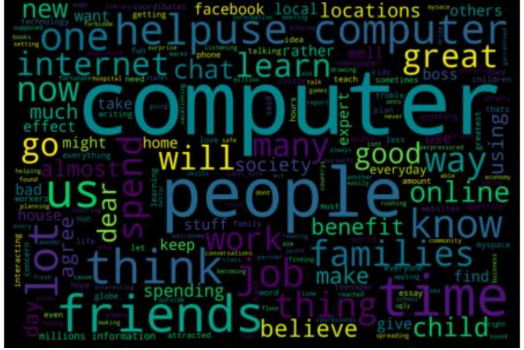
Tianying Luo

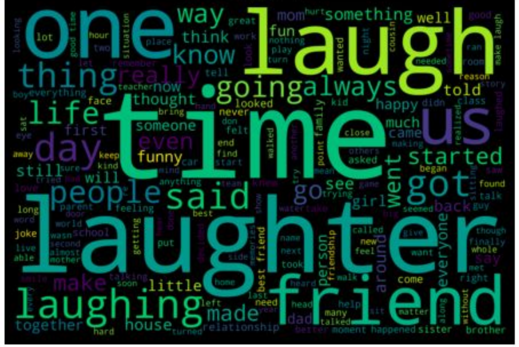
Dataset

Essay Set	Grade Level	Scores Range	Avg Words	Avg Sentences
1	8	2 - 12	366	22
2	10	1 - 6	381	21
3	10	0 - 3	109	7
4	10	0 - 3	94	6
5	8	0 - 4	122	8
6	10	0 - 4	153	9
7	7	2 - 24	168	12
8	10	10 - 60	605	35

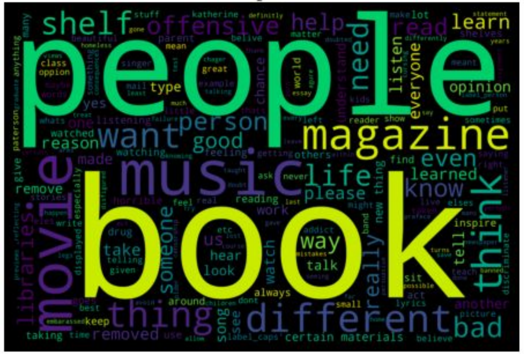
- Kaggle (The Hewlett Foundation)
- 13K Train Essays
- 4K Validation Essays
- 4K Test Essays



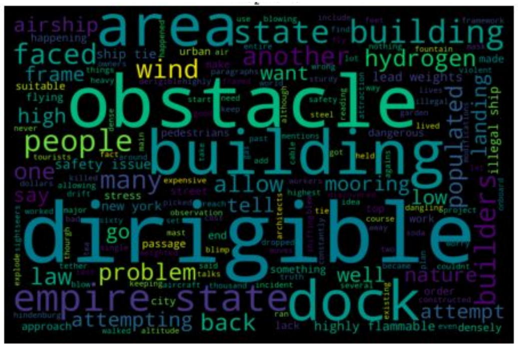

computer
people
families
friends
think
time
will
us
job
know
online
great
internet
learn
chat
helpuse
one

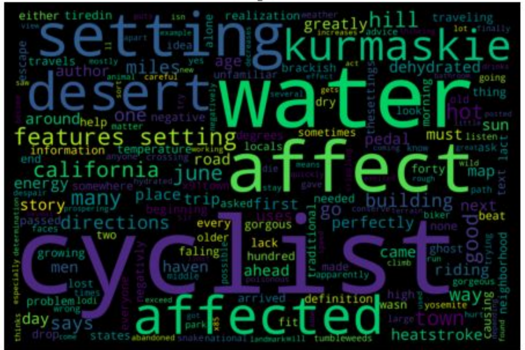

time
laughter
laugh
one
friend
laughing
us
said
got
thing
life
day
people
going
always

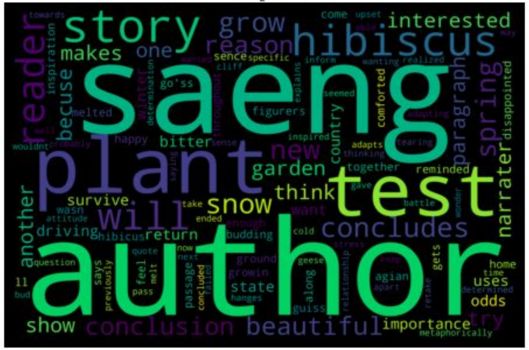

patience
people
sale
lot
mom
job
one
wedding
long
looking
take
patient
time
finally

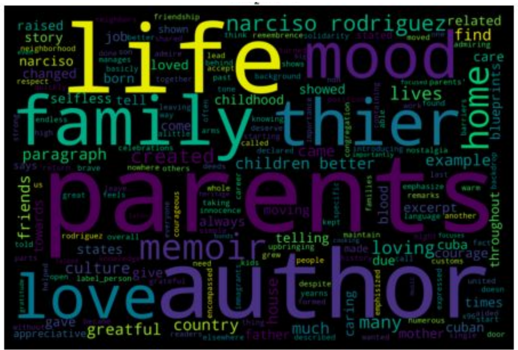

people
book
music
magazine
movie
different
person
want
think
life
shelf

Dataset

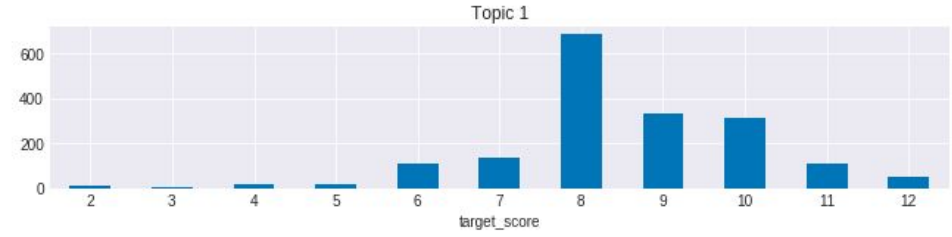

obstacle
building
dirigible
area
state
dock
empire
hydrogen
wind
allow
mooring
people

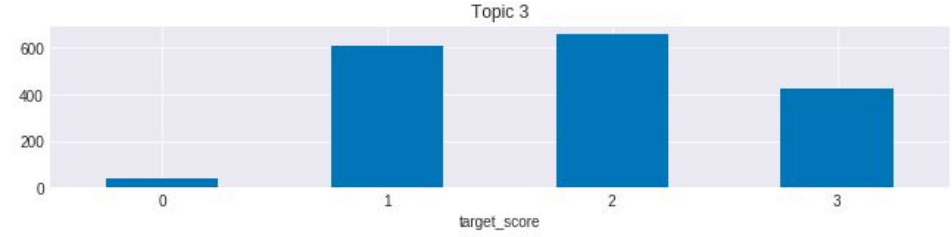

cyclist
affect
affected
water
desert
setting
kurmaskie
features setting
california

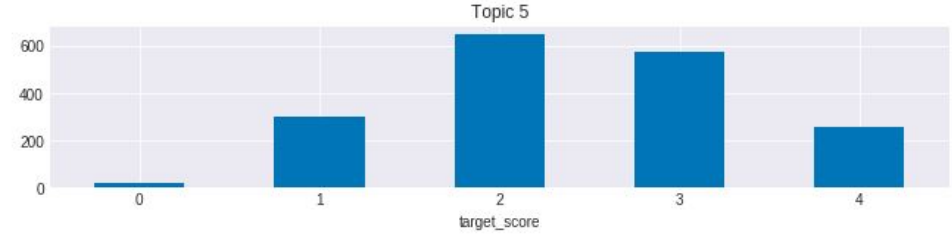

saeng
author
plant
test
story
hibiscus
snow
garden
conclusion
beautiful

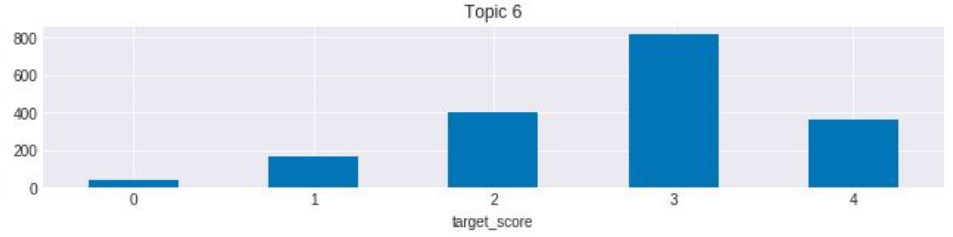

parents
life
family
author
mood
love
thier
memoir
narciso rodriguez

Target Score Distribution - Per Essay Set

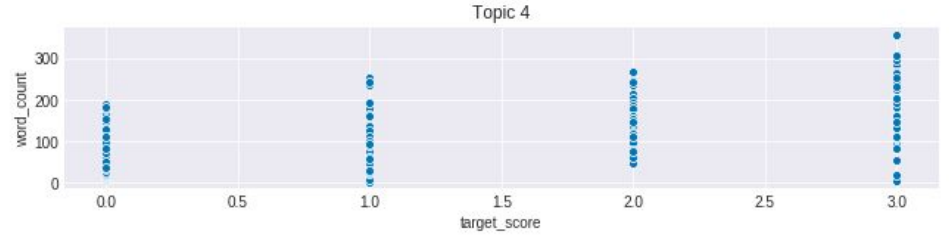


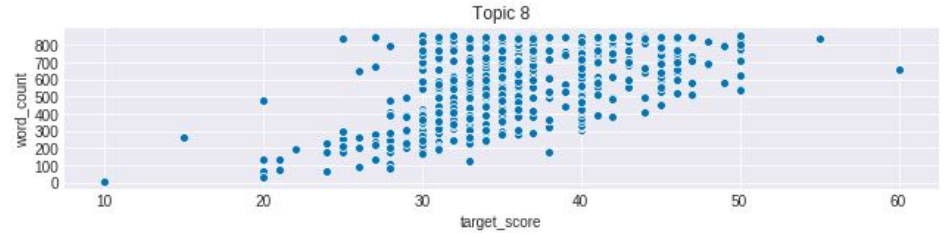






Target Score Distribution - Word Count





Pre-Processing: Essay Corpus

Computers a good because you can get infermation, you can play games, you can get pictures, But when you on the computer you might find something or someone that is bad or is viris. If ther is a vris you might want shut off the computers so it does not get worse. The are websites for kids, like games, there are teen games, there are adult games. Also pictures are bad for kids because most of the time they lead to inapropreit pictures. You should only look up infermation that you need not things like wepons or knifes. Also there are differnt kinds of companies like @CAPS1 @CAPS2. @CAPS2 is a good place to get computers @CAPS1 so is @CAPS1.

Pre-Processing: Essay Corpus

Computers a good because you can get infermation, you can play games, you can get pictures, But when you on the computer you might find something or someone that is bad or is viris. If ther is a vris you might want shut off the computers so it does not get worse. The are websites for kids, like games, there are teen games, there are adult games. Also pictures are bad for kids because most of the time they lead to inapropreit pictures. You should only look up infermation that you need not things like wepons or knifes. Also there are differnt kinds of companies like @CAPS1 @CAPS2. @CAPS2 is a good place to get computers @CAPS1 so is @CAPS1.

computers good get infermation, play games, get pictures, computer might find something someone bad viris. ther vris might want shut computers get worse. websites kids, like games, teen games, adult games. also pictures bad kids time lead inapropreit pictures. look infermation need things like wepons knifes. also differnt kinds companies like label_caps label_caps. label_caps good place get computers label_caps label_caps.

Pre-Processing: Essay Corpus

Computers a good because you can get infermation, you can play games, you can get pictures, But when you on the computer you might find something or someone that is bad or is viris. If ther is a vris you might want shut off the computers so it does not get worse. The are websites for kids, like games, there are teen games, there are adult games. Also pictures are bad for kids because most of the time they lead to inapropreit pictures. You should only look up infermation that you need not things like wepons or knifes. Also there are differnt kinds of companies like @CAPS1 @CAPS2. @CAPS2 is a good place to get computers @CAPS1 so is @CAPS1.

computers good get infermation, play games, get pictures, computer might find something someone bad viris. ther vris might want shut computers get worse. websites kids, like games, teen games, adult games. also pictures bad kids time lead inapropreit pictures. look infermation need things like wepons knifes. also differnt kinds companies like label_caps label_caps. label_caps good place get computers label_caps label_caps.

['computers', 'good', 'get', 'infermation', 'play', 'games', 'get', 'pictures', 'computer', 'might', 'find', 'something', 'someone', 'bad', 'viris', 'ther', 'vris', 'might', 'want', 'shut', 'computers', 'get', 'worse', 'websites', 'kids', 'like', 'games', 'teen', 'games', 'adult', 'games', 'also', 'pictures', 'bad', 'kids', 'time', 'lead', 'inapropreit', 'pictures', 'look', 'infermation', 'need', 'things', 'like', 'wepons', 'knifes', 'also', 'differnt', 'kinds', 'companies', 'like', 'labelcaps', 'labelcaps', 'labelcaps', 'good', 'place', 'get', 'computers', 'labelcaps', 'labelcaps']

Pre-Processing: Essay Scores

Scale Scores to range [0, 1] - per essay set (MinMaxScaler)

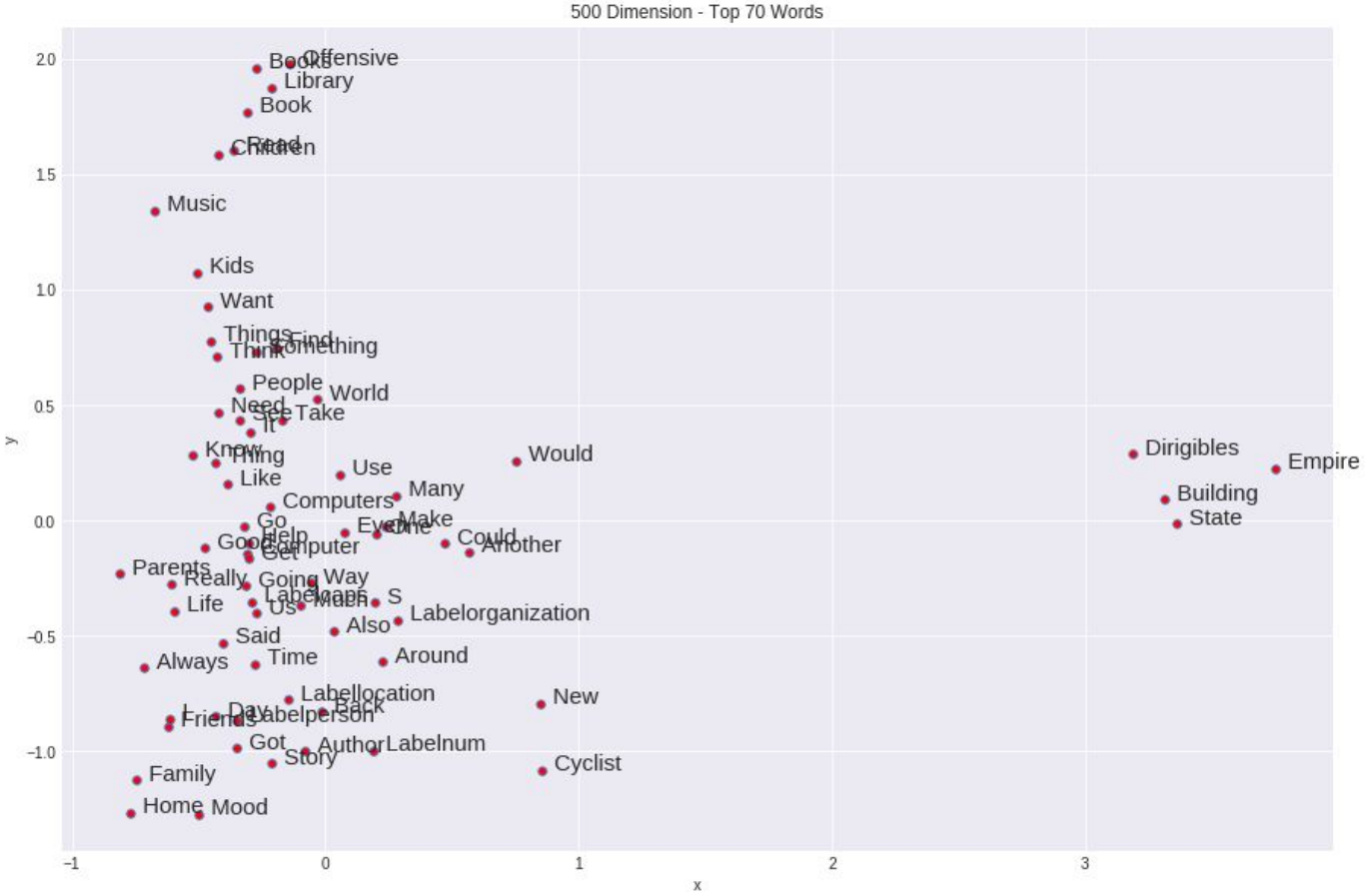
Set 1: [2, 12]

essay_set	target_score	scaled_target_score
1	8	0.6
1	9	0.7
1	7	0.5
1	10	0.8
1	8	0.6

Set 7: [2, 24]

essay_set	target_score	scaled_target_score
7	12	0.454545
7	16	0.636364
7	19	0.772727
7	22	0.909091
7	15	0.590909

Word Embeddings



Word Embeddings

Top 10 Similar Words & Cosine Similarities

‘computer’

computers	0.80
internet	0.73
computor	0.70
compute	0.68
internent	0.67
laptops	0.66
peopel	0.66
electronic	0.66
comuter	0.66
web	0.66

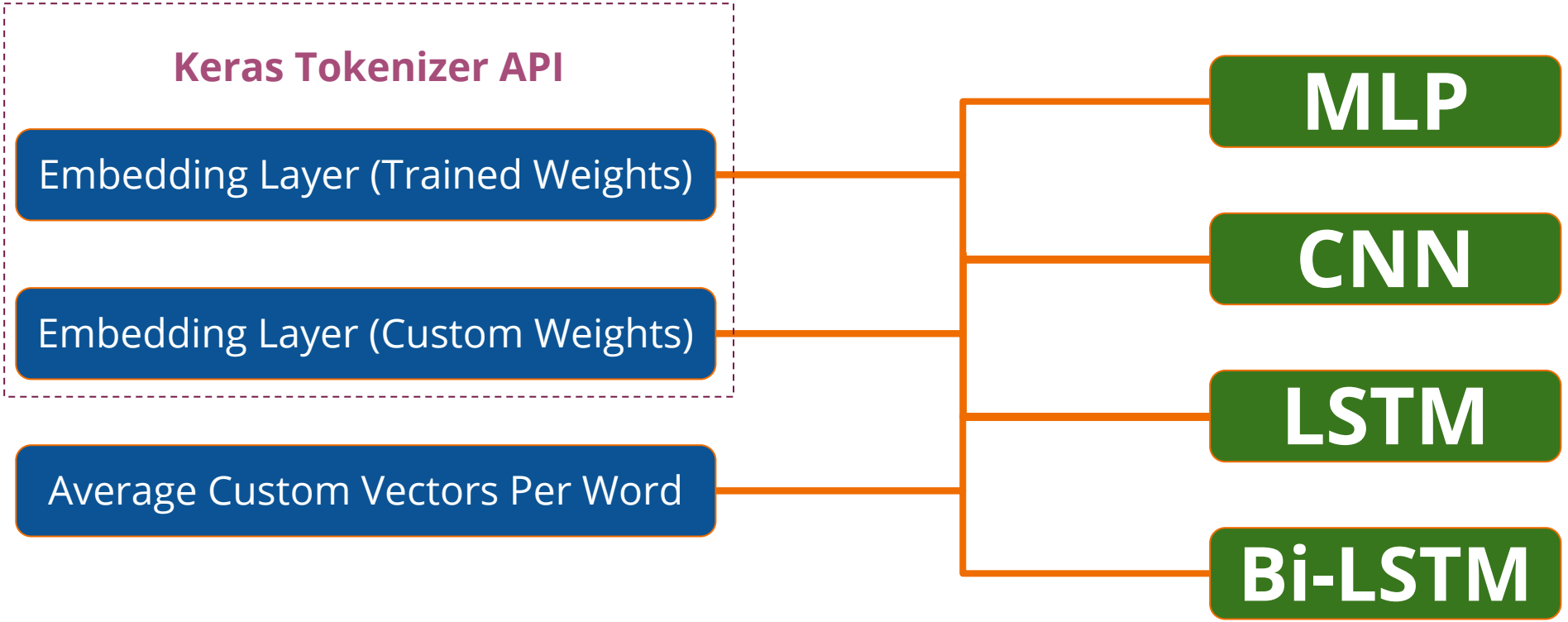
‘laugh’

laughing	0.76
laughs	0.76
joke	0.73
laughter	0.72
ache	0.71
goofy	0.70
giggle	0.70
chuckle	0.68
bust	0.68
obnoxious	0.68

‘cycle’

cycles	0.87
pedaling	0.86
slowing	0.86
hilly	0.86
rocky	0.86
uneven	0.86
uphill	0.86
tiring	0.86
drained	0.85
downhill	0.83

Model Approach



Evaluation Metrics

MSE Loss Function

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Quadratic Cohen Kappa Score

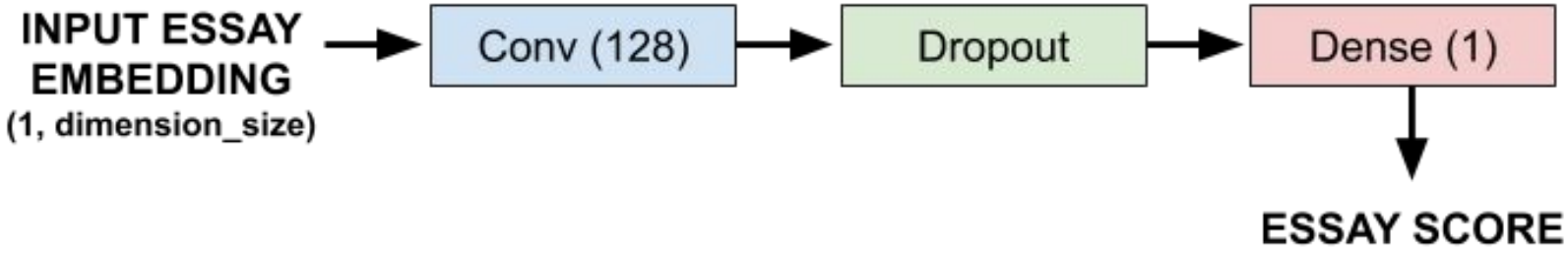
$$\kappa = 1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}x_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}m_{ij}}$$

$$\text{Quadratic: } w_i = 1 - \frac{i^2}{(k-1)^2}$$

Human Raters' κ = 0.7544

Leaderboard κ = 0.81407

Multi-Layer Perceptron

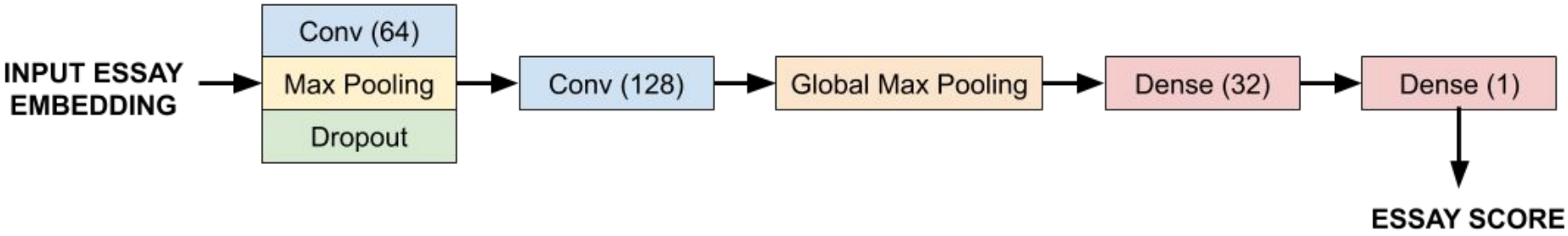


Multi-Layer Perceptron

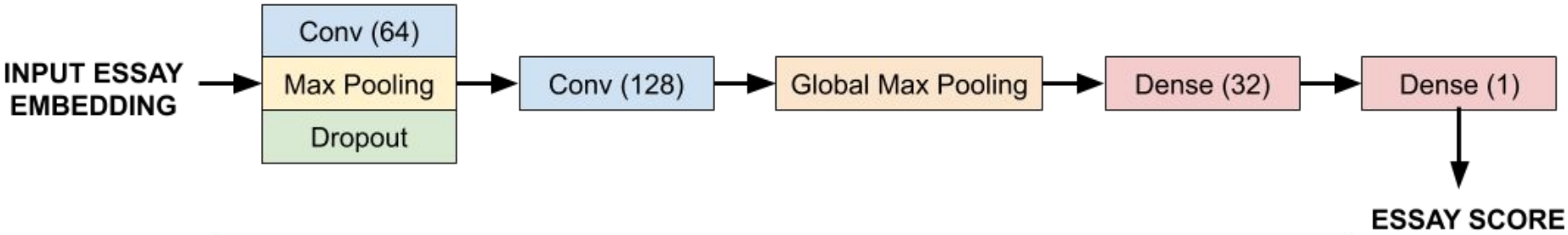
Parameters:
Batch Size = 128, Epochs = 100, Dropout = 0.2, Dimension = 300

Optimization Algorithm	Test Kappa Score
Stochastic Gradient Descent	0.4338
RMSprop	0.6268
Adam	0.6388

Convolutional Neural Network



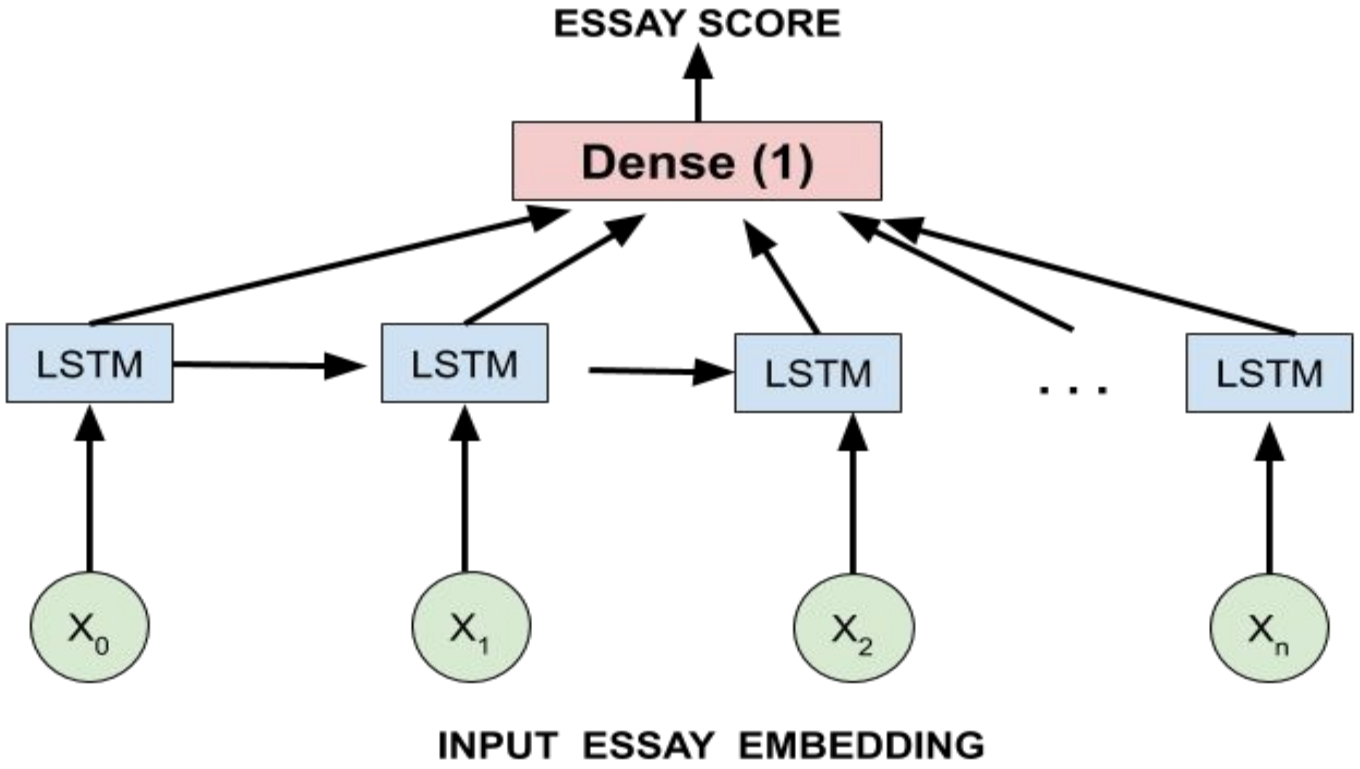
Convolutional Neural Network



Parameters:
Batch Size = 256, Epochs = 100, Adam(lr = 0.001), Dropout = 0.2, Dimension = 100

Embedding Format	Test Kappa Score
Embedding Layer (Trained Weights)	0.5882
Averaged Word2Vec Essay Vectors per Word	0.6237
Embedding Layer (Word2Vec Essay Weights)	0.6808

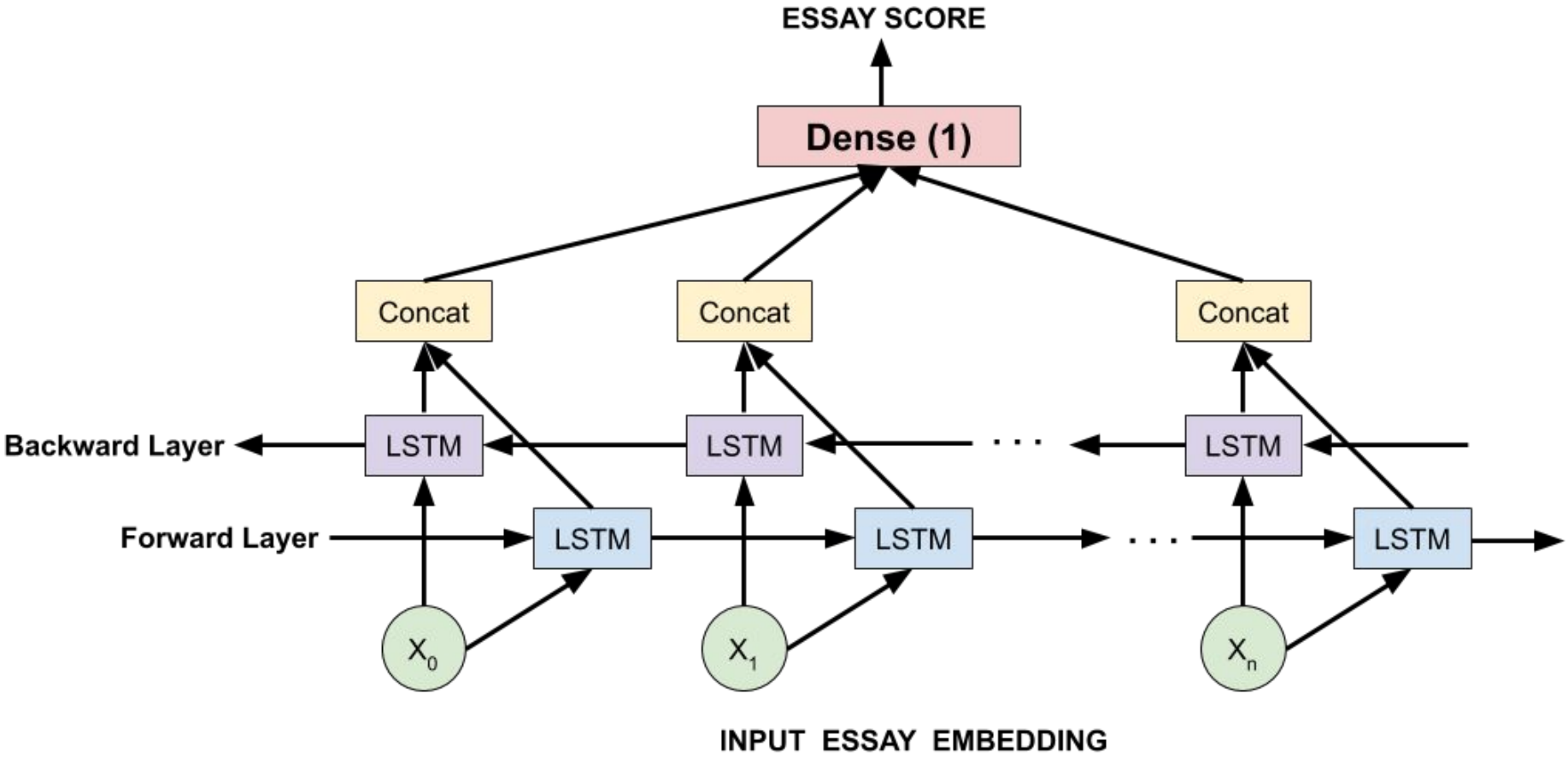
Long-Short Term Memory (LSTM)



Long-Short Term Memory (LSTM)

Parameters: Batch Size = 128, Epochs = 100, Dimension = 100	
Dropout Rate	**Test Kappa Score**
0.1	0.5992
0.2	0.6053
0.3	0.5881
0.4	0.5976
0.5	0.5871

Bidirectional LSTM

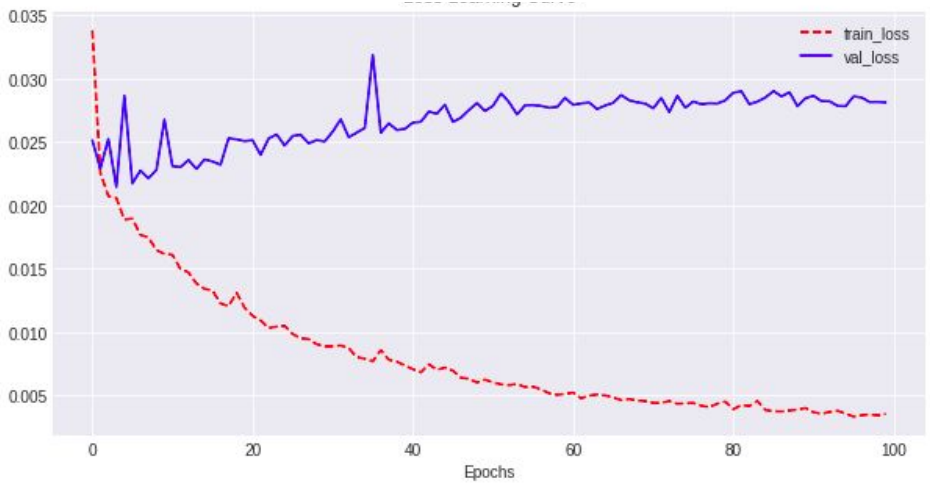


Bidirectional LSTM

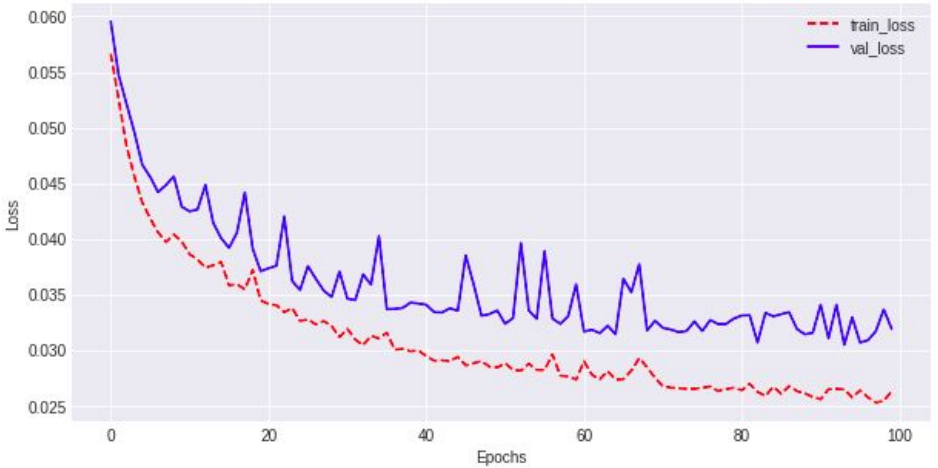
Parameters:
Batch Size = 256, Epochs = 100, Adam(lr = 0.001), Dimension = 100, # Hidden Units = 64

Embedding Format	Train Loss	Val Loss	Test Kappa Score
Embedding Layer (Word2Vec Essay weights)	0.0081	0.0268	0.5299
Averaged Word2Vec Essay Vectors per Word	0.0309	0.036	0.5466

Loss: Embedding Layer Model



Loss: Average Vector Model



Bidirectional LSTM

Essay Set	Scores Range
1	2 - 12
2	1 - 6
3	0 - 3
4	0 - 3
5	0 - 4
6	0 - 4
7	2 - 24
8	10 - 60

Parameters:
Batch Size = 32, Epochs = 100, Adam(lr = 0.001), Dimension = 200, # Hidden Units = 128

Essay Set	Train Loss	Val Loss	Test Kappa	Rater Kappa
1	0.0107	0.0148	0.5934	0.721
1, 2	0.0115	0.0144	0.597	0.765
1, 2, 3	0.0179	0.0243	0.5885	0.7681
1, 2, 3, 4	0.0204	0.0279	0.5795	0.7888
1, 2, 3, 4, 5	0.02	0.0304	0.6007	0.7816
1, 2, 3, 4, 5, 6	0.0193	0.0299	0.6658	0.7807
1, 2, 3, 4, 5, 6, 7	0.0191	0.0297	**0.7043**	0.7723
1, 2, 3, 4, 5, 6, 7, 8	0.0185	0.0283	0.6226	0.7544

Final Results

Model	Test Kappa Score
Multi-Layer Perceptron (MLP)	0.6388
Convolutional Neural Network (CNN)	0.6808
Long-Short Term Memory (LSTM)	0.6185
Bidirectional Long-Short Term Memory (Bi-LSTM)	0.6226

Conclusion & Future Work

- Correct Grammar & Spelling
- Stemming & Lemmatization
- Feature Engineering
- Essay Generation

Thank You!
Questions?