Australian National University

# Explainable Text Classification: Improving lexicon-based classification using n-grams

Supervisors - Priscilla Kan John
Kerry Tylor

COMP4560

By: Jiulin Zong – u6921252

# Background

Learning based approaches for automated text classification tend to be based on complex black-box algorithms that deliver high accuracy predictions but at the cost of not being able to explain the rationale behind their decisions.

In this project, we want to design an experiment to extend a lexicon-based classifier named VADER (Valence Aware Dictionary and sEntiment Reasoner) to investigate how using n-grams will affect its performance. A comparison with learning-based classifiers can also be carried out to find out how performance differs on different types of datasets. By design, VADER has been built to work well on sentiment analysis on social media, it would be of interest to see how VADER (both with unigrams and n-grams) perform in other areas (transfer learning).

# Motivation

- vaderSentiment

```
I like you very much---------------------------------------------- {'neg': 0.0, 'neu': 0.615, 'pos': 0.385, 'compound': 0.3612}
I like you just so so--------------------------------------------- {'neg': 0.0, 'neu': 0.667, 'pos': 0.333, 'compound': 0.3612}
I like you a little----------------------------------------------- {'neg': 0.0, 'neu': 0.615, 'pos': 0.385, 'compound': 0.3612}
I like you-------------------------------------------------------- {'neg': 0.0, 'neu': 0.444, 'pos': 0.556, 'compound': 0.3612}
I hate you just so so--------------------------------------------- {'neg': 0.425, 'neu': 0.575, 'pos': 0.0, 'compound': -0.5719}
I hate you-------------------------------------------------------- {'neg': 0.649, 'neu': 0.351, 'pos': 0.0, 'compound': -0.5719}
I hate you very much---------------------------------------------- {'neg': 0.481, 'neu': 0.519, 'pos': 0.0, 'compound': -0.5719}
I hate you a little----------------------------------------------- {'neg': 0.481, 'neu': 0.519, 'pos': 0.0, 'compound': -0.5719}
AVERAGE SENTIMENT FOR PARAGRAPH:    0.0
```


```
I like you very much---------------------------------------------- {'neg': 0.0, 'neu': 0.615, 'pos': 0.385, 'compound': 0.3612}
I like you just so so--------------------------------------------- {'neg': 0.0, 'neu': 0.667, 'pos': 0.333, 'compound': 0.3612}
I like you a little----------------------------------------------- {'neg': 0.0, 'neu': 0.615, 'pos': 0.385, 'compound': 0.3612}
I like you-------------------------------------------------------- {'neg': 0.0, 'neu': 0.444, 'pos': 0.556, 'compound': 0.3612}
I hate you just so so--------------------------------------------- {'neg': 0.425, 'neu': 0.575, 'pos': 0.0, 'compound': -0.5719}
I hate you-------------------------------------------------------- {'neg': 0.649, 'neu': 0.351, 'pos': 0.0, 'compound': -0.5719}
I hate you very much---------------------------------------------- {'neg': 0.481, 'neu': 0.519, 'pos': 0.0, 'compound': -0.5719}
I hate you a little----------------------------------------------- {'neg': 0.481, 'neu': 0.519, 'pos': 0.0, 'compound': -0.5719}
AVERAGE SENTIMENT FOR PARAGRAPH:    0.0
```

# Motivation

- N-gram

S = ‘I love deep learning’

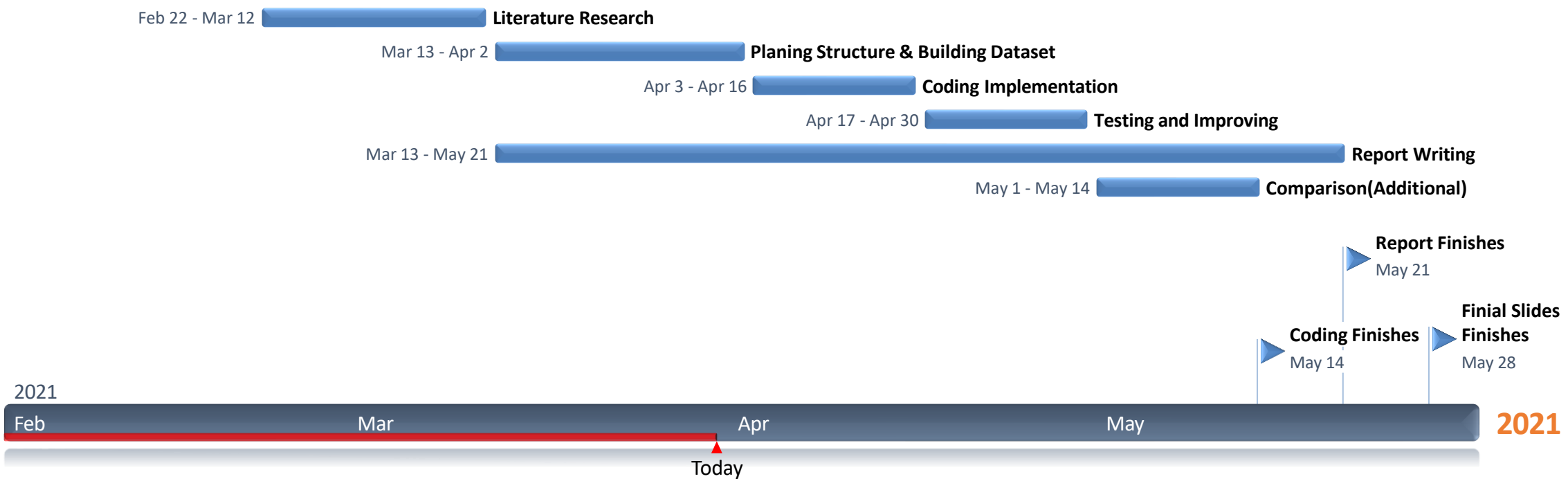
Y1 = {‘I’, ‘love deep’, ‘learning’}
Y2 = {‘I love’, ‘deep’, ‘learning’}
Y3 = {‘I’, ‘love’, ‘deep learning’ }

p(Y1) = p(I)p(love deep | I)p(learning | love deep)
p(Y2) = p(I love)p(deep | I love)p(learning | deep)
P(Y3) = p(I)p(love | I)p(deep learning | love)

# Process

1. Literature research on language models (especially n-grams) and classifiers
2. Building data sets with sentiment ranks
3. Implementation and testing datasets.
4. Comparison.

# Timeline


Feb 22 - Mar 12
Literature Research
Mar 13 - Apr 2
Planing Structure & Building Dataset
Apr 3 - Apr 16
Coding Implementation
Apr 17 - Apr 30
Testing and Improving
Mar 13 - May 21
Report Writing
May 1 - May 14
Comparison(Additional)
Report Finishes
May 21
Finial Slides Finishes
May 28
Coding Finishes
May 14
2021
Feb
Mar
Apr
May
2021
Today

# Potential Problems

1. A convincing dataset with good diversity
2. Reasonable sentiment ranks
3. A suitable language model

# Thanks For Watching

Q & A