

Activity 9: Machine Learning (Part 2)

Activity 9: Machine Learning

Group No: 33**Group Member:**

1. Nipat Chenthanakij 6430215121
2. Natyanin(Teerepat) Chayanuwong 6430178621
3. Praty Benjawan 6531323521
4. Chotpisit Adunsehawatt 6531313221

Part 2: Feature Engineering

1. Go to this [link](#), follow the instructions inside the provided Google Colab notebook. Then, answer the questions below.

1.1 Given a post with 500 shares, which page is it more likely to come from?

dramaaddict

1.2 Given a post published at 20:30, which page is it more likely to come from?

ejeab

1.3 Given a post with 140 characters, which page is it more likely to come from?

ejeab

1.4 Discuss how you come up with the list of words.

Activity 9: Machine Learning (Part 2)

By randomizing a word and checking whether the word percentage is significantly distinct or not

1.5 Write down your final list of words together with their feature importance and the training accuracy. Also include a screenshot of the training result.

List_words = ['หมา', 'แมว', 'coffee', 'กา', 'ไก่', 'สัสน', 'ควย', 'เหี้ย', 'เย็ด']

List words, Thai or English, you want to include in the features of model. (e.g. ['หมู', 'หมา', 'กา', 'ไก่'])

```
##### <YOUR CODE HERE> #####
# Select word feature of model. to increase Test accuracy
word_features = ['หมา', 'แมว', 'coffee', 'กา', 'ไก่', 'สัสน', 'ควย', 'เหี้ย', 'เย็ด']
##### <END OF YOUR CODE> #####
trainWithWords(data, word_features)
```

word_features: ['หมา', 'แมว', 'coffee', 'กา', 'ไก่', 'สัสน', 'ควย', 'เหี้ย', 'เย็ด']
Train Accuracy: 80.44%

	features	feature_importances
1	hour_published	0.438626
0	num_shares	0.419125
2	message_length	0.127492
6	has_กา	0.014757
3	has_หมา	0.000000
4	has_แมว	0.000000
5	has_coffee	0.000000
7	has_ไก่	0.000000
8	has_สัสน	0.000000
9	has_ควย	0.000000
10	has_เหี้ย	0.000000
11	has_เย็ด	0.000000

tree.pdf has been created.

Here's the performance of your model on another data set we have set aside for evaluation:

Test Accuracy: 64.55%

Training accuracy = 80.44%

1.6 Write down your test accuracy, also include a screenshot of the testing result. Discuss how it differs from the training accuracy.

Activity 9: Machine Learning (Part 2)

List words, Thai or English, you want to include in the features of model. (e.g. ['หมี', 'หมา', 'กา', 'ไก่'])

```

✓
0s
▶ ##### <YOUR CODE HERE> #####
# Select word feature of model. to increase Test accuracy
word_features = ['หมา', 'แมว', 'coffee', 'กา', 'ไก่', 'ลีส', 'ควย', 'เหี้ย', 'เย็ด']
##### <END OF YOUR CODE> #####
trainWithWords(data, word_features)

word_features: ['หมา', 'แมว', 'coffee', 'กา', 'ไก่', 'ลีส', 'ควย', 'เหี้ย', 'เย็ด']
Train Accuracy: 80.44%

      features  feature_importances
1  hour_published      0.438626
0   num_shares      0.419125
2  message_length      0.127492
6      has_กา      0.014757
3      has_หมา      0.000000
4      has_แมว      0.000000
5    has_coffee      0.000000
7      has_ไก่      0.000000
8      has_ลีส      0.000000
9      has_ควย      0.000000
10     has_เหี้ย      0.000000
11     has_เย็ด      0.000000

tree.pdf has been created.
Here's the performance of your model on another data set we have set aside for evaluation:

Test Accuracy: 64.55%

```

Test Accuracy = 64.55%

1.7 There will be a “tree.svg” file created in the file directory of Google Colab. Upload it to Google Drive and put the link here.

<https://drive.google.com/file/d/1ATyW8jIECYEoJg1g1Uy1qfCKS0hFuY0Q/view?usp=sharing>

Once you finish, students must inform an instructor or a TA for inspection.

— THIS IS THE END OF PART 2 —